


```

67  easnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnT 84
   |||||||
7912 CAATAAATATGGAACCCAGCAACCAATATCGGATGTTTCCAAACT 7863
   |||||||
84  yraAallePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
   |||||||
7862 ACGCTATTTTCCCATTTGATGCTCCGAGCAACGTTGCTTTGGTCTT 7813
   |||||||
101 MetGlnLysValProLysGluLeuIleGlnGlnThrAsnLysTyr 117
   |||||||
7812 ATGCAAAAGAAGCTTCCAAAAGAAGATTGATTACAGACCAACAAGTA 7763
   |||||||
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysL 134
   |||||||
7762 TCTTGAACTCATGCAAAATGCTCAATATGCGATCGAAAGCCCATTAAC 7713
   |||||||
134 euserGlyGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
   |||||||
7712 TCAGTGGTGGACACAAACGTTGTCACCTTGGCATCGCCCTTAAGCGGTT 7663
   |||||||
151 AspProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaL 167
   |||||||
7662 AATCCAACTGTTCTCTCATGAGACGACCACTTGTATCTGGAGGCCAA 7613
   |||||||
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
   |||||||
7612 ACTTCGCTGGATGCGTCACAGCCATCCGAAATCCACACAGAAATGG 7563
   |||||||
184 LyleThrThrValTyrValThrHisAspGlnGluGluAlaMetAlaIle 200
   |||||||
7562 GAATTAAACCTGTTATGTAAACCCAGCAACGAAGAAGCATGGCTATT 7513
   |||||||
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
   |||||||
7512 TCAGACCAAAATGCTGTTATGAAGAATGGGGTGATCCACAAATCGGCCG 7463
   |||||||
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234
   |||||||
7462 ACCAAAGAAGACTCTATCATTAACCAAGCTAATGAGTTGTGGCAACCTTTA 7413
   |||||||
234 leGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
   |||||||
7412 TCGGACCCCAAAATATTATCCCTGCCAATCTTGAAAAACGAGGAGCGGC 7363
   |||||||
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
   |||||||
7362 GCTTATATCGCTTTTCAGATGGCTATGCCCTTGCAATGCCAGCTTTGA 7313
   |||||||
267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluLubPheI 284
   |||||||
7312 TCAGGTTAGAGAGCAAGCTATTATGTAAGCATTCGTCGCAAGAGTTTA 7263
   |||||||
284 leLysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300
   |||||||
7262 TCAAGAATGATCTGGAGATTATTAAGCAACTATTTAAGATAGCGCTAT 7213
   |||||||
301 leuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317
   |||||||
7212 CTTCGATTAATACGATTTATTTTCATTGAGACAGGTTTTCCTCAAAAAT 7163
   |||||||
317 eGlnValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsnA 334
   |||||||
7162 TCAAGTTAGGAAGAATCAACTTTTGAAGAAGATCTCAAAAAGCAATC 7113
   |||||||
334 rglIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
   |||||||
7112 GTATTGCTCTACGATCAATACGCAAAATTTAAACATCTTTTCGACAGAT 7063
   |||||||
351 gLysSerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363
   |||||||
7062 GGTTCCTCCAAACCTGATATAAAGAGTCAACCATGCAACG 7024

```

```

seq_name: gb_ba:AE007337
seq_documentation_block:
LOCUS      AE007337                10216 bp    DNA        linear    BCT 31-Aug-2001
DEFINITION Streptococcus pneumoniae TIGR4 section 20 of 194 of the complete genome.
ACCESSION  AE007337 AE005672
VERSION    AE007337.1 GI:14971709
KEYWORDS
SOURCE
ORGANISM   Streptococcus pneumoniae TIGR4.
            Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE
  1 (bases 1 to 10216)
    Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
    Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J.,
    Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,D.D.,
    Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
    Holtzapple,E., Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L.,
    McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
    Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
    Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
    Fraser,C.M.
    Complete genome sequence of a virulent isolate of Streptococcus
    pneumoniae. Science. 293 (5529), 498-506 (2001)
  2 (bases 1 to 10216)
    Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
    Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J.,
    Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,D.D.,
    Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
    Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A.,
    Feldblyum,T.V., Angiuoli,S., Gwinn,P., Hickey,E.K., Holt,I.E.,
    Loftus,B.J., Uyal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
    Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
    Direct Submission
    Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
    Medical Center Dr, Rockville, MD 20850, USA
    Location/Qualifiers
      1..10216
        /organism="Streptococcus pneumoniae TIGR4"
        /strain="TIGR4"
        /db_xref="taxon:170187"
      153..845
        /gene="SP0240"
      153..845
        /gene="SP0240"
      153..845
        /gene="SP0240"
        /codon_start=1
        /transl_table=11
        /product="phosphoglycerate mutase family protein"
        /protein_id="AAK74420.1"
        /db_xref="GI:14971710"
        /translation="MKVRLVLRHGKTMFNTIGRAGKSDPTLAREGRIOELGIC
        LRESLQFERAVSSDSGRITDTMGILLLELGLGEPYRMDRIREWPGSDGAYVDG
        DLFGGILIRIFNVHDVHQLSYELAELGAEVDTAAGAWGWEKLSGRIEGEMIAKEM
        EDQGGNALVYVSHGWTIGTIYVLINGMHPHGIDNSVILLEYEDQFVEVYVGDSTR
        ELGRKMEGSI"
        complement(948..2593)
        /gene="SP0241"
        /note="This region contains a gene with one or more
        premature stops or frameshifts, and is not the result of a
        sequencing artifact. This region contains an authentic
        frame shift and is not the result of a sequencing
        artifact.; similar to GP:2766195; identified"
        complement(2583..3674)
        /gene="SP0242"
        complement(2583..3674)
        /gene="SP0242"
        /codon_start=1
        /transl_table=11
        /product="ABC transporter, ATP-binding protein"

```

```

/protein_id="AAK74421.1"
/transl_table=1
/db_xref="GI:14971711"
/translation="MSEIKIINAKKIHYDVPIENLITIPKSLFTLLGAGCGKTT
LLHAGFNSIEGGEFEEDTKINNMPKRNIGMONTAIPPHLVRPNVAGLMO
KYPKEELIOOTNKYILELMOIAYADKPKLSGGOORTYLLACALAVNSVLMDE
LSNLEALRLDMRAIREIQHEWGITVVYVYHDEEMALISDQIAVMKDVLOQIGRP
KELVHPANEVAFATIGRTNIPANLEKRSYGATVSDGALMPALDVEEQAIHV
SIRPEEFIKDESGDIEGTRSDVYLGINTDYFIETGASKIQVSEESTFEEDLOKGNR
ILRIINTOKINPESGSONLIGVNHGT"
complement(3687. .4703)
/gene="Sp0243"
complement(3687. .4703)
/gene="Sp0243"
/codon_start=1
/transl_table=1
/product="iron ABC transporter, iron-binding protein"
/protein_id="AAK74422.1"
/db_xref="GI:14971712"
/translation="MKRKMVYAAACSSNESADSSDKDGSILVYSPNSEGLIGAT
IPAEERYGIKVELIOAGTGFELFKLESEKEVPADVIFGGSYQYTTGELFENYTS
KENDNVIREONTGTSGPTYLGGSVLIINPDILKGMNIESYNDLFRPELKGATAD
PANSSAPAOITNMLQAGCKYDKDKANSYKDELFTLLDKIGSSSSSVYKVDGEVA
VGLSEDPAAKLMDGANIKVIFKEGVTELPASAIYKSKNMENAKKLTDFIISGE
VQDTLGTITNRPYRNNAKTSNNKPIDKIKTIEDYVIKKNKSDIVKKNVEFTDI
QSKO"
complement(5008. .5115)
/gene="Sp0244"
complement(5008. .5115)
/gene="Sp0244"
/identified by glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK74423.1"
/db_xref="GI:14971713"
/translation="MSHSFKSLQKEILHRSSIAAPVTSRAFSDFVSP1"
complement(5290. .6066)
/gene="Sp0245"
complement(5290. .6066)
/gene="Sp0245"
/identified by match to Pfam protein family HMM
PF02143"
/codon_start=1
/transl_table=1
/product="pyruvate formate-lyase-activating enzyme,
putative"
/protein_id="AAK74424.1"
/db_xref="GI:14971714"
/translation="MEISGIIIFNIQHSIHDPGIRTVFLKGCPLRCPWCSNPESO
RMPEKKKADQREKFTLVGEKEVTEITTEVLKDKFEYEESSGGTLLSGGEIFQOFER
AKILTKSAKEHHITAIETPAVDHEKFTIDLOYVDFTYDLKHNYSNKKKKYGVFN
OMITKNHFAFSQNKTYLRLPVLPFNNSNLEDEKERTLPLNSLINDQVLLPFRHQS
ENKRYRLNRRKEMGINALHPEDLDYQKVLNHHINCYF"
6188. .6934
/gene="Sp0246"
6188. .6934
/gene="Sp0246"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, Deor family"
/protein_id="AAK74425.1"
/db_xref="GI:14971715"
/translation="MKRLKIOIKLVSEYEKIDVNTSKLNVSVTTRKDKLKSNG
LLHREHCAYANSGDDLNVRLSINYEKRIKYQAVKLYSNEMIMESSTCALLAE
ETCKOKRWITVINSFPLANFVRAVDSCRIVLGEFGQDSQVTVGLPKMTIDTFHY
QCAFVGIDYDKEMGFTGKDLRSEVVOYISAVDKYIVLIDSSKFKRGTVRRFALS
QVVEVITDEKLSKONIAITLENAGIMVYVS"
6947. .7927
/gene="Sp0247"
6947. .7927
/gene="Sp0247"
/identified by sequence similarity; putative"

```

```

/codon_start=1
/transl_table=1
/product="transcriptional regulator"
/protein_id="AAK74426.1"
/db_xref="GI:14971716"
/translation="MNDRKRLSKTAVLYIENLNSQIAKGIYRISISMTEA
RMVGIETIEENDTNMFKLENVYKEYSLESELEIPNEDDPTTILSERISQVAVG
LNLVIDNMKEGYSWGSLSNLVDLHRSKSVRNHFFPLAGPSHIAKYHVNLLYE
MSRKFEGECTFMATVYQENKLLADGILQSRFENLKNMSKMDIAVVGIGDPSNGK
HQMIDMTEDDEKELTKVYKVGELCCFPKSKREYENQERTIASLSDLANIPOS
LAVAGDTKVSILVRLANLVNHLITDKNTILKLEDEGLTFREITGE"
8122. .8442
/gene="Sp0248"
8122. .8442
/gene="Sp0248"
/codon_start=1
/transl_table=1
/product="PTS system, IIA component"
/protein_id="AAK74427.1"
/db_xref="GI:14971717"
/translation="MEMIVPDOIIMGLILYAGDAKOHYKALDYIKNGTCEGEEIO
LADALIEAHNLOTKFLAQASGKTEITLAFVHSQDHLMTSMTEINLKEITISKE
LHKR"
8488. .8796
/gene="Sp0249"
8488. .8796
/gene="Sp0249"
/codon_start=1
/transl_table=1
/product="PTS system, IIB component"
/protein_id="AAK74428.1"
/db_xref="GI:14971718"
/translation="MVKIGLFCAGFSTGMLVNNMKIAAGSGVEATEEAFSOKIAD
YAPNIDVALLGPQVAYRLDKSKELCDKCDVPIAVIPMDGMGMDGKVLDAISLISG
"
8789. .10111
/gene="Sp0250"
8789. .10111
/gene="Sp0250"
/codon_start=1
/transl_table=1
/product="PTS system, IIC component"
/protein_id="AAK74429.1"
/db_xref="GI:14971719"
/translation="MKRRKRSIMKNVOKIAPMKFVNVRGIIAKDKMALTPT
VGSFLIMGOLPEGLNKSTIASVFGNMEPEPMOYSGFPAIMGLISCSIASYAK
NSGVEALPAGLVSAFPIILRSSYIPKQGBAIGDAISKVFGGQIIGAIITGLVG
SIYFFIKRKIVIMPEQVQAIKQEFAMI PAVIFLSSIVYILAKSLTNGTFLIE
MYSALIOPLQGLTSGYGAIGAFISFLMWFVHGOSVNGVYVTLALLSNDPANKA
MLASANLSLENGAHIVQOFLDSFLILSSGSIPTGLVYVAMLPFAKSKOYALGVAAF
PAIFNVNPEVYFGPIVMPNPMVPEFIIYVLAIVYGYALAGFPMQPSGVILPMST
PAISGLVGMQGVITQVILAMSLTYVFPFFKVDRLAYQNEIKOS"
BASE COUNT 3252 a 1519 c 2088 g 3357 t
ORIGIN
alignment_scores:
Quality: 1843.00 Length: 363
Ratio: 5.077 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment block:
US-09-769-787-162 x AE007337/rev ..
Align seg 1/1 to reverse of: AE007337 from: 1 to: 10216
1 MetSerGluIleIleIleAsnAlaIleTyHisAspValPr 17
|||||
3674 ATGAGTGAGATCAAAATTTATTAACGCCCAAAAATCTACACAGATGCTCC 3625
17 oValIleGluAsnLeuAsnIleThrIleProIysGlySerLeuPheThrL 34
|||||
3624 TGTATTGAGAAATTTGACATTTACAAAGAGAGACTCTTATACC 3575

```

34 euleuglyalaSerGlyCysGlyThrThrLeuLeuArgMetIleAla 50
 3574 TTCTGGAGCTTCAGAGATGTGGAAAAAGACCCTTCTCGATGATGCA 3525
 51 GlypheasnsertilegilyglygluPhethepshaspsapthrysl 67
 3524 GGTTCACAGCTATCGAAGGTGGAGAAATTTACTTCAGATGATACAAAAT 3475
 67 easnasmetgluproserlysarasnileglymetvalpheglnasnt 84
 3474 CAATAAATATGGAAACCCAGCAACCAATATGGGATGTTTCCAAACT 3425
 84 yralailepheprohisleuthrvalargspasnavalaalpheglyleu 100
 3424 ACGTATTTTCCCATTTGACTGTCCGAGACAACGTTGCTTTGGCTT 3375
 101 Metglnyslysva1ProlysgluLeuileglnIntnrranlysty 117
 3374 ATGCAGAAAGAGGTTCCAAAGAAAGAAATTAATTCACAGACCAAGTA 3325
 117 rleuglulemetglnileagIntnrra1aasparglysproslysl 134
 3324 TCTTGAACCTCATGCAATTCCTCAATATGGGATCGAAAGCCGATTAAC 3275
 134 euSerGlyglynglnargvalthrleualaacyalaaleuala1a1 150
 3274 TCAATGTGGACACACAAACGTCACCTGGCATCGCGCTTAGCGGT 3225
 151 AsnproservalleuLeuMetaspgluproLeuSerlsnleuglualy 167
 3224 AATCCAAAGTTCCTCCATGAGAGACCCACTAGTAATCTGGAGGCCAA 3175
 167 sleargleuaspmetargrlna1a1learglileglnhislgluvalg 184
 3174 ACTTCGCTTGATATGCGTCAGAGCATCCAGAAATCCACACAGAAATGG 3125
 184 lytlethrthrvaltyrvalthrHisaspnglnuglualametalale 200
 3124 GAATTAACACTGTTATGTAAACCACGACCAAGAAAGCAATGGCTATT 3075
 201 SeraspnglnleualaValMetlysaspglyvalileglnIntnleegy 217
 3074 TCGAACCAAAATGCTGTTATGAAGATGGGTGATCCAAACAAATGGCGC 3025
 217 gProlysgluleuTyrrHislysproalaasngluphevala1athrphel 234
 3024 ACCAAAGAACTCATATCAACACAGCTAATGAGTTTGGCAACCTTTA 2975
 234 leglyargthrasnileleproalaasnleuglulysarqseraspgly 250
 2974 TCGAGACCAAAATATATATCCCTGCCAATCTTGAACCAAGCAGCAGCG 2925
 251 AlatyrrilevalpheseraspglytyrAlaLeuargmetproalaLeu 267
 2924 GCTTATATCTCTTTTCAGATGGCTATGCCCTTCGAATGCGACGCTTTGA 2875
 267 pglInvalgluglna1a1leHisvalserileargProglugluphe1 284
 2874 TCAGGTTGAGAGACCAACCTATTCATCAACATTCGCCGGAAGAGTTTA 2825
 284 lelysaaspgluserglyaspIleuglythrIleargaspservaltyr 300
 2824 TCAAGATGATCTGAGATATGAAGAGAACTATTAGAGATAGCGTCTAT 2775
 302 leuglyleuasnthraspytyrrheileglnthrglyphealeaserlysl 317
 2774 CTTGGAGTAAATACGGATTAATTCATTGAGACAGGTTTCCCAAAAT 2725
 317 eglInvalserglugluserthrphleglugaaspleugllyslYAsna 334
 2724 TCAAGTTAGTGAAGAAATCAACTTTTGAAGAAGATCTTACAAAAGGCAATC 2675

334 rg1leargleuarglileasnthrGlnlysleuasnillepheserAlaasp 350
 2674 GTATTCCTTACGAAATCAATACGCAAAATTAACATCTTTCTGCAGAT 2625
 351 GlySerGlnasnleuilelysglyvalAsnHISglythr 363
 2624 GGTTCACCAAACTGATTAAGAGATCAACCATGGAACG 2586

seq_name: gb_ba:AE008403

seg documentation block:

LOCUS AE008403 9976 bp DNA linear BCT 13-SEP-2001

DEFINITION Streptococcus pneumoniae R6 section 19 of 184 of the complete genome.

ACCESSION AE008403 AE007317
 VERSION AE008403.1 GI:15457762

KEYWORDS

SOURCE

ORGANISM

Streptococcus pneumoniae R6.
 Streptococcus pneumoniae R6
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 (bases 1 to 9976)
 Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
 Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C.,
 Gilmore, R., Glass, J.S., Khoja, H., Kraft, A., Ladace, R.,
 Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
 Mcahren, S., McHenry, M., Mclester, K., Mundy, C., Nicas, T.I.,
 Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
 Sun, P.-M., Winkler, M.E., Yang, Y., Young, Bellido, M., Zhao, G.,
 Zook, C., Baltz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L.
 and Glass, J.I.
 Genome of the Bacterium Streptococcus pneumoniae strain R6
 J. Bacteriol. 183 (19), 5709-5717 (2001)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 11544234

REFERENCE
 2 (bases 1 to 9976)
 Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
 Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C.,
 Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Ladace, R.,
 Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
 Mcahren, S., McHenry, M., Mclester, K., Mundy, C., Nicas, T.I.,
 Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
 Sun, P.-M., Winkler, M.E., Yang, Y., Young, Bellido, M., Zhao, G.,
 Zook, C., Baltz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L.
 and Glass, J.I.
 Direct Submission
 Submitted (27-JUN-2001) Infectious Diseases Research, Eli Lilly and
 Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA

FEATURES
 Location/Qualifiers
 1..9976
 /organism="Streptococcus pneumoniae R6"
 /strain="R6"
 /db_xref="taxon:171101"

gene
 117..476
 /gene="spr0217"

CDS
 117..476
 /gene="spr0217"

note="ACT domain containing protein"

/codon_start=1
 /transl_table=11

/product="Conserved hypothetical protein"

/protein_id="AAK99021.1"

/db_xref="GI:15457763"

/translation="MSIFVRIMLYIKRIENYVSYFKKMSILMKAITTVGKDS
 GIVAGVSKIAELGINDIDISQTVLDEFTFMAYVSSDEKODFTYLNREPARQOTLN
 VKINQSAIPEAMYN1"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

486..1823
 /gene="spr0218"

/label=spr0218
 /product="conserved hypothetical protein"
 /protein_id="AAK9022.1"
 /db_xref="gi:15457764"
 /translation="MDIROVETETAMIEONFDIRITTMGISLDCIDPNNINRAEKI
 YOKITTKVANIIVAGDEIAEIAELGPIYKKRVSVPISLIGAAPATQYVLAALADKA
 AKETITVDITGFSALVQCKYOKGDEIILNSTIPALAEIDKCSSVNIIGTSKSNMTRA
 VADMERITIKETANLSDMGVAKLVFPAANVEDNPPMAGFHVGGADVIIVNGVSGPV
 VKRALEKESGSDVVAETVTKAFKTIIRIQQLGOMASERLGEVFIIVDSLAPTPA
 VGDVARYLVEELGSETVGTGHTTAAALNIOVKRGKGMACNOYGGSLGAFIPSEDEE
 GMAIVANGSNLLEKLEAMTAICSGIDMIIPDTPPETITAAIADGAIVGNMKT
 TAVRIIPGREGDEMTIEFGILLGTAPVMKVNQASSVDFTISGQILPRTIHSKN"
 2010..2759
 /gene="gpmB"
 /note="spr0219"
 2010..2759
 /gene="gpmB"
 /EC_number="5.4.2.1"
 /codon_start=1
 /transl_table=11
 /label=spr0219
 /product="Phosphoglycerate mutase 2 paralog"
 /protein_id="AAK9023.1"
 /db_xref="gi:15457765"
 /translation="MGGFYFIIHDIIEANGSRKWKVRLIYRHKTFMENTIGNAGSM
 SDPTIFRAGEIDELGIGRESIDQFERAYSDSGRTIQMGIIIEELGQGTIPR
 MDKIRREMGFSGDAYSVDGLFMGIIPRIENVDHVHOLSVALEGLVEDYDAWAGG
 LEYEGSRIRKEGFEMIAKEMEDOGGNALVYSHGMTIGTIYLYINGMHPHGDINSVTI
 LEYEDGQPRVEVYAGRSYRELGRKMEBGS"
 complement(2862)..3896
 /gene="ABC-MSP"
 /note="spr0220"
 complement(2862)..3896
 /gene="ABC-MSP"
 /note="Putative"
 /codon_start=1
 /transl_table=11
 /label=spr0220
 /product="ABC transporter membrane-spanning permease -
 iron transport"
 /protein_id="AAK9024.1"
 /db_xref="gi:15457766"
 /translation="MLVFRAFSDFGTIPMLIGEGRTFPPVIITYQFISENGNSAFAI
 MAIIIALIFLQKHIANRFAFSNMLLHPIPKRTTKGMAIYATVYGIIFTSVDPQ
 ILYLTSFTLKTSGMFEFGYSPNSYKLAFNMGSAIENTIRIPLALVVLVLPFTFS
 YLAVERKMLNFELIDLSIMVPYIYVGTVLGIAPTSFNTGLFGSGPLMTGTAILIM
 SLAKRLYTIIRSSVASIQOIAPISEEAESIGSSRLNIPAKITTPMMLISITGAILLS
 SMVYIISELSITLILYNKTRMYAITYEVLKRGYGAALSTILVLYVGSLLFPM
 KTSKNSITL"
 complement(3881)..4507
 /gene="ABC-MSP"
 /note="spr0221"
 complement(3881)..4507
 /gene="ABC-MSP"
 /note="Putative"
 /codon_start=1
 /transl_table=11
 /label=spr0221
 /product="ABC transporter membrane-spanning permease -
 iron transport"
 /protein_id="AAK9025.1"
 /db_xref="gi:15457767"
 /translation="MERKKLIMTASFEFLFLYLVLPVPIITVLKOALIHESQSL
 ANFVPEFVAGVSWLLRGNVITKEFLNALHLRAIDYIGKGIYVPIQLPFLVFL
 VYAGTMSINDSLSEPAESMGSGFGKPIVTVPLVPLVTLAACLIL"
 complement(4497)..5588
 /gene="ABC-NBD"
 /note="spr0222"
 complement(4497)..5588
 /gene="ABC-NBD"
 /note="Putative"
 /codon_start=1

```

/transl_table=11
/label-spr0222
/product="ABC transporter ATP-binding protein - iron
transport"
/protein_id="AAK99026.1"
/db_xref="GI:15457768"
/translation="MSIEIILNKKITHDVPIENLITIPKSLFTLLASGCGRTT
LKRNIAGFNSIKDGEYFDPIKINMPPSKRNIGLVQNTAIFPHLTVRONVAGLMQ
KVKPEELIOOTNRYLYELMOIAQVADRPKLSGGQOQRYTLACALVNSVLLMDP
LSNKEALRLDMQNTIIEIHEVEGIVTVYVYHDEEAMASDQIAVAKDGVIOIGRP
KEIYKHPANEFVATFIRGTNIPANLEKRSQGAIVPSDGAALMPALDDOVBQAIHV
SIRREIFKDESGDIEGTIRDSYGLNTDPTFETGFASKIQVSEESTFPEELQKGNR
IRLRINTQKLNIPEADGSONLIKGVNHT"
complement(5601..5969)
/gene="ABC-SBP-truncation"
/note="spr0223"
complement(5601..5969)
/gene="ABC-SBP-truncation"
/note="Putative"
/codon_start=1
/transl_table=11
/product="ABC transporter solute-binding protein - iron
transport_truncation"
/protein_id="AAK99027.1"
/db_xref="GI:15457769"
/translation="MTVGLSYEDSAVKLLNDGANIKRVYPREGTVLPASAAIVKSK
NMEAKKFIDFIISOEVQDTLTGTTTNRPPVKNKATSENMKPIDKIKTLEDYDYVIK
NKSIVKYKNEVFVDIOSQ"
6020..6126
/rpt_family="RUP element"
/rpt_type=dispersed
/rpt_unit=Rup-A
complement(6114..6308)
/gene="ABC-SBP-truncation"
/note="spr0224"
complement(6114..6308)
/gene="ABC-SBP-truncation"
/note="Putative"
/codon_start=1
/transl_table=11
/label-spr0224
/product="ABC transporter substrate-binding protein - iron
transport_truncation"
/protein_id="AAK99028.1"
/db_xref="GI:15457770"
/translation="MNIIEGSDLLKPELKGKIAIADPANSASSAPQLTNMQAGGYK
DKKMSIVKDLFTLIDSVK"
6799..6897
/rpt_family="RUP element"
/rpt_type=dispersed
/rpt_unit=Rup-B1
complement(6840..7137)
/gene="spr0225"
complement(6840..7137)
/gene="spr0225"
/codon_start=1
/transl_table=11
/label-spr0225
/product="Hypothetical protein"
/protein_id="AAK99029.1"
/db_xref="GI:15457771"
/translation="MTLNEYILQYRLKQALDKMAESPNSPLSAISDQVGSYKYPFAK
VFKKYHISPKKLSGLRIYK"
complement(7212..7988)

```

Align seg 1/1 to reverse of: AE008403 from: 1 to: 9976

```

1 MetSerGluIleLysIleIleAsnAlaLysLysIleTyrHisAspValPr 17
|||||
5588 ATGAGTGGATGCGAATATTATTAACGCCCAAAATAATCTACCGAGATGTC 5539
17 ovalleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
|||||
5538 TGTATTGAGATTGAGACATTACCAATTCCAAAGAGAGTCTCTTACC 5489
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
|||||
5488 TTCCTGGAGCTTCAGAGATGGGAAACGACCTCTTCGATGATGATCA 5439
51 GlyPheAsnSerIleGluGlyGlyLysPheTyrPheAspPheThrLysI 67
|||||
5438 GGTTCACACAGATCAAAAGATGAGAAATTTCCTCGATGATACAAAAT 5389
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84
|||||
5388 CAATGATATGAGAACCCAGCAAAATATCGGCTGTTTCCAAACT 5339
84 YrAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
|||||
5338 ACGCTATTTCCTCCACATTGACTGTCCAGACAACTGCTTTGGCTCT 5289
101 MetGlnLysValProLysGluLeuIleGlnIleThrAsnLysTyr 117
|||||
5288 ATGCCAAAGAGGTTCCAAAAGAAATGATTTCACAGACCAACACATGA 5239
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLys 134
|||||
5238 TCTTGAACCTCATGCAAAATTCCTCATATGCGGATCGAAGCCGATAAC 5189
134 euseGlyGlyGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaVal 150
|||||
5188 TCAGTGGGAGCAACACACGCTGCACCTGGCTGCGCTTACGGCTT 5139
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAla 167
|||||
5138 AATCCAAATGTTCTCTCATGAGACGCACTTGAATCTGGAGGCCAA 5089
167 sLeuArgLeuAspMetArgIleAlaIleArgGluIleGlnHisGluValG 184
|||||
5088 ACTTCGCTTGATATGCTCAAAACATCCGAGAAATCCAAACGAGATGG 5039
184 LylIleThrThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200
|||||
5038 GAATTCAACACTGTTATGTAAACCAACGACCAAGAAAGCATGGCTATT 4989
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyArg 217
|||||
4988 TCAGACCAAAATTCGCTTATGAAGATGGGTGATCCAAACAATCGCGCG 4939
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||
4938 ACCAAAGAACTCTATCATTAACACAGCTAATGAGTTTGGCAACTTTA 4889
234 leGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
|||||
4888 TCGGACGCGCAAAATATTATCCCTGCCAATCTTGAAGAAAGGAGCGACG 4839
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
|||||
4838 GCTTATATTCGCTTTTCAGATGGCTATGCCCTTCGAATGCCAGCTTTGA 4789
267 pGluValGluGluAlaIleHisValSerIleArgProGluGluPhe 284
|||||
4788 TCAGGTTGAGGAGCAAGCTATTTCATGTAAGCATTCGTCGCAAGATTTA 4739
284 leLysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300
|||||

```

```

4738 TCAAAGATGAATCTGAGATATCGAAGCACTATTAGATAGCGCTTAT 4689
301 LeuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317
|||||
4688 CTGGACTTAATACGATATTATTTCTAGACAGGTTTGGCTCAAAAAT 4639
317 eGlnValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 334
|||||
4638 TCAAGTTAGTAGAATCACTTTGGAAGAGAACTACCAAAAGCAATC 4589
334 rGlyIleArgLeuAlaArgIleAsnThrGlnLysLeuAsnIlePheSerAla 350
|||||
4588 GTATTGCTTACGATCAATACGCAAAATAATTAACATCTTTCTGCAGAT 4539
351 GlySerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363
|||||
4538 GGTTCGCCAAACCTGATATAAGAGACTCAACCATGGAACG 4500

```

seq_name: gb_hlg:SPNEU1908

seq_documentation_block:

LOCUS SPNEU1908 175936 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION AL449930
VERSION AL449930.1 GI:11545155
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.

REFERENCE 1 (bases 1 to 175936)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 175936)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peltsch,M. and
Garcia-Bustos,J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN

COMMENT
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers

1..175936
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"

BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others

alignment_scores:

Quality: 1807.00 Length: 363
Ratio: 5.006 Gaps: 0
Percent Similarity: 99.449 Percent Identity: 97.796

alignment_block:

US-09-769-787-162 x SPNEU1908/rev ..

Align seg 1/1 to reverse of: SPNEU1908 from: 1 to: 175936

1 MetSerGluIleLysIleIleAsnAlaLysLysIleTyrHisAspValPr 17

```

|||||:|||||
69746 ATGAGTGAATCGAAATTTAAGCCAAAATACTACCATGTC 69697
47 ovalleGIuAsnLeuAnlleThrIleProLysGlySerLeuPheThrI 34
|||||
69696 TGTATTGAGAAATTTGACATTTCAATTCAAAAGAGATCTGTTTACC 69647
34 euleuGIuAlaSerGIuGlyLysThrThreueuArGmetIleAla 50
|||||
69646 TTCTTGAGACTTCAGATGTGGAAAACGACCTTCTGCTATGATGCA 69597
51 GlyPheAsnSerIleGIuGIuGlyPheTyrPheAspSerThrLysI 67
|||||
69596 GGTTCACAGATCAAGATGAGAAATTTACTGATGATCAAAAAAT 69547
67 eAsnAsnMetGIuProSerLysArgAnlleGIuMetValPheGlnAsp 84
|||||
69546 CAAATATATGAGACCAACGCAATATGCGGTGTTTCCAAAAC 69497
84 yAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
|||||
69496 ACCCTTTTCCACATTTGACGTCCGAGACACAGTTGTTGTTGCTT 69447
101 MetGIuLysLysValProLysGIuLeuIleGlnThrAnLysTyr 117
|||||
69446 ATGCAAAAGAGGTTCCAAAAGAAATGATTCACACAGCAACAAGTA 69397
117 rleuGIuLysMetGlnIleAlaGlnPyrAlaAspArgLysProAspLys 134
|||||
69396 TCTTGACATGCAAAATGCTCAATATGGGATGCAAAAGCCGATTAAC 69347
134 euserGIuGlyGlnGlnArgValThrleuAlaCysAlaIleuAlaVal 150
|||||
69346 TCAGTGTGTGACACACAACAGTGTCACTTGCGATGCGCTTAAGCGTT 69297
151 AsnProSerValleuLysMetAspGIuProLysSerAsnLeuGlnAla 167
|||||
69296 AATCCAAAGTGTCTCTCATGAGAGCCACTAGTAATCTGAGGCCAA 69247
167 sleuAlaGlyLeuAspMetArgGlnAlaIleArgGluIleGlnHisGlyVal 184
|||||
69246 ACTTCCCTTGAGATGCGTCACAGCCATCCGAGAAATCCAAACAGAGTGG 69197
184 ylleThrThrValTyrValThrHisAspGlnGIuAlaMetAlaIle 200
|||||
69196 GAATTCACACTGTTTATGTAACCCAGCACCAAGACAGCATGCTATT 69147
201 SerAspGlnIleAlaValMetLysAspGIuValIleGlnGlnIleGlyAr 217
|||||
69146 TCAGACCAATTCGTGTATGAAAGATGGGTGATCCACAAATCGGCCG 69097
217 gProLysGIuLeuTyrHisLysProAlaAsnGIuPheValAlaThrPheI 234
|||||
69096 ACCAAAAGAACTCTATCATTAACACCTAATGATTTGGGCAACCTTTA 69047
234 leGIuYrGthrAsnIleIleProAlaAsnLeuGIuLysArgSerAspGI 250
|||||
69046 TCGGACGCCAAATATATTCCTACTAATCTTAAAAAACGAGCGACGCG 68997
251 AlaTyrIleValPheSerAspGIuTyrAlaLeuArgMetProAlaLeuAs 267
|||||
68996 GCTTATATCGTCTTTTCAGATGCTATGCCCTTCGATGCGACTTCTGA 68947
267 pGIuValGIuGIuGlnAlaIleHisValSerIleArgProGIuGlnPheI 284
|||||
68946 TCAGGTGGAGGACAGCATTCATGTAGCATTCGTCGGAAGAGCTTTA 68897
284 leLysAspGIuSerGIuAspIleGIuGIuTyrThrIleArgAspSerValTyr 300
|||||
68896 TCAAAATGATATCTGAGATATTGAAGGAACCTATTAGCCATACGCTAT 68847
301 leuGIuLysAsnThrAspTyrPheIleGIuThrGIuPheAlaSerLysI 317
|||||

```

```

68846 CTTGGACTAATAATACGAGATATTCATTGAGACAGGTTTGCCTCAAAAAT 68797
317 eGIuValSerGIuGlySerThrPheGIuGIuAsnLeuGIuLysGIuAsnA 334
|||||
68796 TCAAGTTAGTACAGATTCACACTTTTGACAGAGATCTTACAAAAGGCAATC 68747
334 rGIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
|||||
68746 GTATTGCTTCAGCAATCATACGCAAAAATTAAACATCTTTCTGAGAT 68697
351 GlySerGIuAsnLeuIleLysGIuValAsnHisGIuThr 363
|||||
68696 GGTTCACAAAACCTGATTAAGAGATCAACATGGAACG 68658
seq_name: gp_ba:SHU75349
seq_documentation_block:
LOCUS SHU75349 6200 bp DNA linear BCT 18-NOV-1999
DEFINITION Serpulina hydrosentariae bit operon, complete sequence.
ACCESSION U75349
VERSION U75349.3 GI:6127224
KEYWORDS
SOURCE
ORGANISM
Brachyspira hyodysenteriae.
Brachyspira hyodysenteriae
Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
1 (bases 1 to 6200)
Dugourd,D., Martin,C., Rioux,C.R., Jacques,M. and Harel,J.
Characterization of a periplasmic ATP-binding cassette iron import
system of Brachyspira (Serpulina) hyodysenteriae
J. Bacteriol. 181 (22), 6948-6957 (1999)
2 (bases 1084 to 6183)
Dugourd,D.F., Jacques,M. and Harel,J.
Direct Submission
Submitted (17-OCT-1996) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
3 (bases 601 to 6200)
Dugourd,D.F., Jacques,M. and Harel,J.
Direct Submission
Submitted (05-MAR-1999) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
REMARK
Sequence update by submitter
4 (bases 1 to 6200)
Dugourd,D.F.
Direct Submission
Submitted (27-OCT-1999) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
REMARK
Sequence update by submitter
On Oct 27, 1999 this sequence version replaced gi:4337124.
COMMENT
FEATURES
source
1..6200
/organism="Brachyspira hyodysenteriae"
/db_xref="taxon:159"
119..6140
/gene="bit operon"
119..1129
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="periplasmic-iron-binding protein Bita"
/protein_id="AADI8094.2"
/db_xref="gi:6127225"
translation="MKNPILIRCMILSMTLFYSQSSGDSKMANSLVYCSHPDLMMT
ILDFPKKNPDIINEVYTAGGELIKRVARAKMNPIDGVLMGTLNSYKSKTFDENY
TSTNEANTIDFKNTBGPFRFSAPISLIAMNTNLAKIKIEGEDILNPKRTIAA
ADPSASSAFENLVMILYAMKGPDEKWDVQKCAULDKLISGSAYKGVADGS
YVGLIYEPEGISYSSGSPVKVLYMKEGVISPQGVYIIRGKNLKNNAKFKIDYCVS
LDAQNMVLVKLSRSDAIVTDMVKPMSRIVSYITDNADVESRQKMDKFDIPT
SI"

```



```

CDS
    /gene="Atu4581"
    /complement(5273..5803)
    /gene="Atu4581"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAL45375.1"
    /db_xref="GI:17743071"
    /translation="MVNHAHILPLKTSPTISPREKDAAGLAFTVAATASHAGG
    IGRITSMPLISAEPILMISTIDASSRVLDIASRRFSVAALSRGGEIADYAPGGR
    NLPDEFANGSDWVPSSGSPRIAAALLSMDCELVSDIAADHILFVGAHIIATDPLRE
    VLMSERNKGAKEASR"
    /complement(5848..6183)
    /gene="Atu4582"
    /complement(5848..6183)
    /gene="Atu4582"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="ferredoxin"
    /protein_id="AAL45376.1"
    /db_xref="GI:17743072"
    /translation="MPYVTENCIACKRYMDCVEVCPECFEGENMLVIHPDCCIDCG
    ICERISPAALIRPDEAGLHVMDLNRHYSIGWPRHQKRTPPDNADTNMGAAKFSI
    LSKNPGAGS"
    /complement(6288..6887)
    /gene="godb"
    /note="Atu4583"
    /complement(6288..6887)
    /gene="godb"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="superoxide dismutase [Fe]"
    /protein_id="AAL45377.1"
    /db_xref="GI:17743073"
    /translation="MAFQPLDPLPYTHALAPSGMSEETIKFPHDHLHKAVVDMLNKAI
    TETEMKSLIEDIYRGITYEKGAVAQSGIFNNASQNRHNNLFWEIMGFQNIATISPLED
    ALTQSFGEIVALEKQNFALAGASQFSGGAWLVPTDGKLIKITTEGVNPLCFGQAL
    LGGDWEHSHYVLEDEETKRPAYENFLDMLVMOAVARL"
    /complement(7052..7621)
    /gene="Atu4584"
    /complement(7052..7621)
    /gene="Atu4584"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator, TetR family"
    /protein_id="AAL45378.1"
    /db_xref="GI:17743074"
    /translation="MPRRKSIISDQVLDALPLMLQTPGGLFFAAAKAGLSAATL
    VQRYKRDLDIVETVILRAMDRLRAETETADEEPLTSCGAILSLRLMPOSSEONAS
    DGLLRDLIRNPKLRAGALMGHTLHAIGRLASDPKDRERLGMQMAVWGCAHTW
    WAFVGEAGASLRRALEWVAVVNVER"
    /complement(8195..8195)
    /gene="Atu4585"
    /complement(8195..8195)
    /gene="Atu4585"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAL45379.1"
    /db_xref="GI:17743075"
    /translation="MAVIRGRIASLDGQIVSGSGSDMLFKKDDMDLGHMDRLFLK
    GRTVYVNRATYIDFLTASQSPWVVGDDQVIVVTSRPLEERGPGLLEVINDVGSILREIR
    ALDDVWVKLGGGKIQMAFLRWGALDEIEIYIPEILIGSEPELPLTGHASPTLVSA

```

```

gene      TAIEKGVRLHWRS"
           complement(8281..9348)
           /gene="Atu4586"
CDS       complement(8281..9348)
           /gene="Atu4586"
           /note="identified by sequence similarity; putative; ORF
           located using Blastx/Glimmer"

alignment_scores:
  Quality: 731.50      Length: 382
  Ratio: 2.729        Gaps: 8
  Percent Similarity: 70.157      Percent Identity: 42.408

alignment_block:
  US-09-769-787-162 x AE009386 ..

Align seg 1/1 to: AE009386 from: 1 to: 10356

1 MetSerCululellystIlelleasAlaLysLysIleYrHisAspValPr 17
||||| :||||| :||| :||||| :|||||
1277 ATGGCCCGCCGAGAAATACAGCATCAAAAGATTCACCGGGATGTCGT 1326

17 ovalleGluAsnLeuAsnIleThrIleProlYSGLSerLeuPheTrL 34
||||| :||||| :||||| :||||| :|||||
1327 TGCCTCTCCGACATCAATATCTCATTCATTCAGATGCTTTTTCACGC 1376

34 euLeuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
||||| :||||| :||||| :||||| :|||||
1377 TTTCTTGACCCAGCGGTTGTGGCAAGACACAGCTTGTGGCAGCATGGCC 1426

51 GlyPheAsnSerIleGluGlyGlyPheThrPheAspAspThrIle 67
||||| :||||| :||||| :||||| :|||||
1427 GGTTCACACGACGACGACGCGCATGACATCGAAAGACAGCGCAT 1476

67 eAsnAspMetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84
||||| :||||| :||||| :||||| :|||||
1477 AGACACAGTGCCTGCCACAGCCGATGTCGGCATGCTGTTTCACAGATT 1526

84 yralIalIePheProIleLeuThrValAlaArgAspAsnValAlaPheGlyLeu 100
||||| :||||| :||||| :||||| :|||||
1527 ATGCAGTCTTTCCGATATTTCCGTATTCGACAAACATTCGTTTCGGCTGC 1576

101 MetGlnLysLysValProlYSGLuGluLeuIleGlnIleThrAsnLys 117
||||| :||||| :||||| :||||| :|||||
1577 AACACGCGCAAGTCATCTTCAGCCGAGATCGCGCAAGCGGTGGCAGAT 1626

117 rLeuGluLeuMetGlnIleAlaGlnYrAlaAspArgLysProlAspLys 134
||||| :||||| :||||| :||||| :|||||
1627 TCTTGAGTGTGTCGACGCTGCTCCCTCTCTCAACGCGATTCGCGACAC 1676

134 euSerGlyGlnGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaVal 150
||||| :||||| :||||| :||||| :|||||
1677 TCTCGCGGCGGACAGCACAGCTGTGTGTAGCGCGGACAGCTGTATC 1726

151 AsnProSerValLeuLeuMetAspGluProlLeuSerAsnLeuGlnAla 167
||||| :||||| :||||| :||||| :|||||
1727 AACCCCAAGTCTGTGATGAGCAACCCGCTGCCAACCTTGACGCCAA 1776

167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184
||||| :||||| :||||| :||||| :|||||
1777 ACTTTCGGGTGGATTGGCGGGAACATGCGGTGAATCCACAGGCCATGA 1826

184 LyIleThrThrValYrValYrHisAspGlnGlnAlaMetAlaIle 200
||||| :||||| :||||| :||||| :|||||
1827 ACATCACACACCGTCTATGTCAAGCATATACGAGAAGGGCGTTGCCATG 1876

201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyArg 217
||||| :||||| :||||| :||||| :|||||
1877 TCCGATCTGTTGTCGTCATGTATGCGGCGTCATTACAGCAGCGTCAC 1926

217 gProlYSGLuLeuYrHisLysProlAlaAsnGluPheValAlaThrPhe 234
||||| :||||| :||||| :||||| :|||||
1927 GCCGTGGGAAGTCTATATCAATCCCGCAACCGCTTTGGTGGCTCCCTGC 1976

```

```

234  leglyargthrAsnIlelleProAlaSnleuGluLyArgSerAspGly 250
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1977  TGGAGGACCAATATTCCTCAGCGTCGAC.....CCGACGGGTGGG 2017
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
251  AlATyrlleValPheSerAspGlyTyrlAlaLeuArgMetProAlaLeuAs 267
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2018  CAACATCATCATCCAGC.....CTCAGGTGTGACGCTGCCGCGACGGCGA 2061
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
267  pGlnVal...GluGluGlnAlaIleHisValSerIleArgProGluGluP 283
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2062  TGGGATTTCCAGGACCGAGCATGTGCTGCGATCCGCCGCGACGCGA 2111
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
283  lle.....LysAspGluSerGlyAspIleGlu 292
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2112  TTTCATATGCTCCTTCATAGGCAATGTGATGAGCCATCGAATCGCC 2161
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
293  GlyThrIleArgAspSerValTyrlLeuGly.....LeuAsnThrAspTy 307
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2162  GTGACCATCAGCGACGTCAGCTTCACTGTCGTGATGAAACGTGCGCGC 2211
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
307  rPheIleGluThrGlyPheAlaSerLys..... 316
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2212  CGTGCCTCTCGTCGGGTAGGAATCGAGCGCATCCAGACCATCGCGG 2261
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
317  .....IleGlnValSerGluSerThrPheGlu..... 326
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2262  AGATCATCGCGCTTCAGCCAAACCAAGAAACGACGTTCTCTCGTTGGC 2311
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
327  .....GluAspLeuGlnLysGlyAsnArgIle..... 335
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2312  GCGCATACGCAATGTTCGGGTGATACCGCGGAGCGCGCTGTCATG 2361
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
336  .ArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2362  ACGGATATGGCTGCGACTGCACAGCAGCGGGCGTTTCGCGTCGAC 2407
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
seq_name: gb_ba:AE008229
seq_documentation block:
LOCUS      AE008229          11697 bp    DNA          linear      BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 33 of
            187 of the complete sequence.
ACCESSION   AE008229 AE007870
VERSION     AE008229.1 GI:15158623
KEYWORDS
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
            Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
ORGANISM    1 (bases 1 to 11697)
            Hinkle,G., Slater,S.C. and Goodner,B.
            Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
            Unpublished
JOURNAL     2 (bases 1 to 11697)
            Hinkle,G., Slater,S.C. and Goodner,B.
            Direct Submission
            Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES
            Location/Qualifiers
                source
                    1..11697
                        /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                        /strain="C58"
                        /db_xref="taxon:181661"
                        complement(70..900)
                        /gene="AGR_L_572"
                        complement(70..900)
                        /gene="AGR_L_572"
                        /note="hypothetical protein"

```

```

/codon_start=1
/transl_table=11
/product="AGR_L_572p"
/protein_id="AAK88858.1"
/db_xref="GI:15158624"
/transl_table="MLKFSAPFAVAWAFSATPAAGADQVAATPAGVGDGPMFOATGYL
MATGANGINISPEORAPGLTIVEKFSFSDVDDLFQGLNIGWGYDFYVSGDMYVDTT
DSKAGLNPALOIPLPTLPACATIDANVDQEPATLOAGYRIIDADGFSLDLAGA
RPMHTSNDVAWASHLSIGVSATHTRESGWDVPVIGARAPESLFTDKLSVSGVDVIGG
FCAGSDLTMSALATAYNYIFTDTLSAGKYVDVYSHDGYVFDLSRSGPYGLTYRF
"
1009..2076
/gene="AGR_L_573"
1009..2076
/gene="AGR_L_573"
/note="probable hydrolase YCA0877 [imported] - Vibrio
cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_L_573p"
/protein_id="AAK88859.1"
/db_xref="GI:15158625"
/transl_table="MRPKTLPEFSACTAIESIAITLTSVAEACTREYVIGENNOVMT
ARSDMKTDVGTNLWVFPKMERSGEAGPNSVKWTSKYGSYIASGYDVSTDGNNBAG
LAANYLMVESYSPYDGKSPGLSIAAMQYLDNFATVEEAVYLEKNPFIITVDSV
PGEERLATLHLSDASGDSATVEYIDGQVTHHROXOVMTNSPTPEOLANAYWT
OIGCTVM.PGTNRASDRYRASFEYANAIIPKSNPEALIASVPSYRMTSVPGITTPD
OPNISTRRRTYIDHKKLTYFESALTFPWFIDMTKIDLSKETGAVKKDLGANOIH
IYSGANESLKDTKPKFPLGL"
complement(2162..2689)
/gene="AGR_L_575"
complement(2162..2689)
/gene="AGR_L_575"
/note="conserved hypothetical protein PA3469 [imported] -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_575p"
/protein_id="AAK88860.1"
/db_xref="GI:15158626"
/transl_table="MAVIRGYTAASLDQIYSGDGLSLMLFKYDDMDLGEHDYRLF
GIRTVMGNAITDFIASQSPWVYGDQRYIVTSPRIEPPKGLPVYNDVGLIELR
ALDDGDVMLGGKLOMAFLERGAIDELIYIYIPILSGEPLFLTGHRASPTLVA
TAIEKGVMHLHRFS"
2736..3305
/gene="AGR_L_578"
2736..3305
/gene="AGR_L_578"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_578p"
/protein_id="AAK88861.1"
/db_xref="GI:15158627"
/transl_table="MPRHSISDDOVLDPALPLMLTQGPGLTFAAACAAGLSAATL
VORAGREDIVETVTLRAMDRSLRAETGDEDEEPILISGASLILRIAMPPOSABONAS
DGLLLRDIRNPKIARAGALMGHTLAHLGKRLASDADRRLGMQAAVVOGAGHTW
WAFVRGEDGSAIRRLERWAVAVVNDER"
3347..4069
/gene="AGR_L_579"
3347..4069
/gene="AGR_L_579"
/note="(AF022931) superoxide dismutase"
/codon_start=1
/transl_table=11
/product="AGR_L_579p"
/protein_id="AAK88862.1"
/db_xref="GI:15158628"
/transl_table="MDLPKRNLTSTVILEVLSPCPKALSMILSLDITKEYMAF
QUPDLPTHTDALAPSGMSEETLKPHHDLHKKAYVNUNLKAITGTREBKSSLDLYRGT
YKGAVAOQGITFNNASQHNWNLFWELKPOHNAIPSLDELATLQSFGLVAFKONFA
LAGASQFGSGMMLVVDVTDGKLTIRTEGVNPLCGOKALLGCDVWHSYIIDRTK
RAYVENFLDNLVNWQAVARL"

```


gene 4162..4509
/gene="AGR_L_582"
CDS 4162..4509
/gene="AGR_L_582"
/note="(AF228062) ferredoxin II"
/codon_start=1
/transl_table=11
/product="AGR_L_582p"
/protein_id="AK88863.1"
/db_xref="GI:15158629"
/translation="MSIPMPYVVTENCIAKYMDCVEVCEPGEYEGENMLVIHPDOC
IDGICERECPAAIRPDPAGLHVMLDNRHSGLMPRHQKRTPPDNADITMNGAAG
KESILSKNPGAGG"
4521..5084
/gene="AGR_L_583"
4521..5084
/gene="AGR_L_583"
/note="(MADH:FMN OXIDOREDUCTASE)"
/codon_start=1
/transl_table=11
/product="AGR_L_583p"
/protein_id="AK88864.1"
/db_xref="GI:15158630"
/translation="MSHNKRLIDGCVNNNAHIFPKTSTPSTISPREKQAMAGLFTV
AAVATSHAGERIGRTITSEPLSAEPPLMI SIDASSRMVDLIAARRRSVVALSRGO
EEIADVPAGKGNLPRDRAVGSMDVWPSGSRFLAALLSMDCELVGSIDADHILFVGA
IIEATCDPLREVLIMSESRNYKGAASR"
complement(5275..6231)
/gene="AGR_L_584"
complement(5275..6231)
CDS /gene="AGR_L_584"
/note="(hypothetical protein)"
/codon_start=1
/transl_table=11
/product="AGR_L_584p"
/protein_id="AK88865.1"
/db_xref="GI:15158631"
/translation="MNIASFQSKRYNNAATPGDDVALILRGVAFVFDGADTPTGANY
DGLSGRIALAAANAVNAEMSIKGTLCNCESTLFPOTISDILKEAERINASHPST
LAIVADMGSLRLLAGDSGIRINGDRLOHSLIDIRVTSARIHVFKLOFAFGSD
AAAKTRATITFSGLRRAAVAKIMKGYDAEKVITTAAGSLQMEITAVEEFLMGIRV
OPHYANRGHVLTGASINGMWEKSTVDLSAKDVTNSLEITDGLSLPTGIDVDD
WEAEIARVETADHKLEYPVSKGTSSEFSDDRTVYCMKAR"
complement(6313..7998)
/gene="AGR_L_586"
complement(6313..7998)
CDS /gene="AGR_L_586"
/note="(586"
/codon_start=1
/transl_table=11
/product="AGR_L_586p"
/protein_id="AK88866.1"
/db_xref="GI:15158632"
/translation="MTDMAATATGGRFLRPGFMTSVVFAVILAIPLIPESV
FISFDKATGAEFSNVAEFTREFTVALMNTLVIGMGLACILGLPFACTISRY
RIKGRFIATFALVLCAPFEGIAVIMLFGNGAVTNLISFAGISLPIVIGIPIV
MYSELEFPYIYLMENTALNTINKSYDAENLGCYAFRFRITLPLVEPAVSTGAI
ICFVLSIADGEPFALIGKIRLSTIAVOYTESMAGPTMAVTISMVIAISMAALL
LORHILAKRYASISLTNPVKOSMRPIKSFIVGRCYIVAFITMISLTVIYISPLAT
SGPYTGGGGLDSYKRILDSFOALNSFMFLAAVYLAIVSGLSFLIVRDNVAVS
GLIDLAMPYILPGVVAIGVYTRPEPDITGALLIILVLFIRLPLVYGRSTTS
ILROIPISEBAVNAGSPAKTFLSVTVPLMLPGLIISLMSFIRAIHELSTLLIY
TARTIMPVLIVIOVIDGEFGTAALSTLLVSTGCVFVERLSDBKASFY"
complement(7995..9119)
/gene="AGR_L_589"
complement(7995..9119)
CDS /gene="AGR_L_589"
/note="(U75349) putative ABC transporter Btld"
/codon_start=1
/transl_table=11
/product="AGR_L_589p"
/protein_id="AK88867.1"

/db_xref="GI:15158633"
/translation="WVACRAGYVMEQHMAVEITSIKSYRDVVALSDINISIPSGSF
FTLIGSGCGKTLTLTLAGFHOQDGSISIERQALIEHVPAHKRDGVAFODVAVP
ISVNDNIAGFKOKRSSSAIEIRVVKIIDVVOALYAKRMPHEISGGQOQVGLARA
LVINPKVILMDEPLSNLDARKLRRLREIIOQANITTVYTHQOEALAMSLDVC
VMVGVIQOAPAPMEVYNNPAPNFVA SPGANNEFLDRTGQSTISIKVTLPOADA
ISKDRITVAIRPEALISTGSPGNCDERIELPYTITVOSTFGRENNVAVLSGEEIE
ATKPSPEITAILOPNQKTFSPWFGDTOLFQDGVTERLS"
complement(9192..10355)
/gene="AGR_L_591"
complement(9192..10355)
CDS /gene="AGR_L_591"
/note="(U75349) periplasmic iron-binding protein Bita"
/codon_start=1
/transl_table=11
/product="AGR_L_591p"
/protein_id="AK88868.1"
/db_xref="GI:15158634"
/translation="MFRITVHFNFNEFALDLILYKKINSRQRAVTTKNSKISRG
TSCVRQKNGAPFMRQSCYIFALVAGATFLASAAFAHAGETVYVYSAAPQDLMDLL
PLFEKKTGKVELVKAGSGELMNRIRIKETGAAGDVIMSVDTVIDFSADLEPYKPV

alignment_scores:
Quality: 731.50 Length: 382
Ratio: 2.729 Gaps: 8
Percent Similarity: 70.157 Percent Identity: 42.408

alignment_block:
US-09-769-787-162 x AE008229/rev ..

Align seg 1/1 to reverse of: AE008229 from: 1 to: 11697

1 MetSerGIuileIySlelleAsnAlaIySlySleIyRhIsaPvalPr 17
||||| :||||| :||| :|||||
9080 ATGGCGCCGCTAGAAATTACACGATCAAGAAAGTTACCGGGATGTCGT 9031
17 ovalIleGIuansleuAsnIlethIleProlyGlySerleuPetherL 34
:||||| :||||| :||||| :||||| :|||||
9030 TGCCTCTCCACATCAATATCTCCATTCATCCAGGATCGTTTTCACGC 8981
34 euleuGIAlaSerGIySGlySyrThrIleuArGMeIlela 50
||||| :||||| :||||| :||||| :|||||
8980 TTCCTGGACCCAGCGGTTGTGGACACGCTTGGCGACGATGCC 8931
51 GlyPheAsnSerIleGIuGIyGIuPetherYrPheAspSyrThryI 67
||||| :||||| :||||| :||||| :|||||
8930 GGTTCATCCACAGACGACGCGCATCATCAAGAACAGCGCAT 8881
67 eAsnAsnMetGIuProSerIySyrGsnIleGIyMetValPheGIuAsnT 84
:||||| :||||| :||||| :||||| :|||||
8880 AGAGCAGCTGCTGCCCCACAAAGCGCATGCGCATGCTTTCACAGATT 8831
84 YrAlaIlePheProHIsleuThrValArgAspAsnValAlaPheGIuYeu 100
||||| :||||| :||||| :||||| :|||||
8830 ATGCAGTCTTCCGATATTTCCGATTCGACAAACATGCTTGGCTCG 8781
101 MetGIuIySlyValProlySGluGIuLeuIleGIuInthrasnIyS 117
||||| :||||| :||||| :||||| :|||||
8780 AACACGCCAGATCATCTACCGAGATCCGCGACGGCGGCAAGAT 8731
117 rLeuGIuLeuMetGIuIleAlaGIuIyAlaAspArgIySProAspIyS 134
:||||| :||||| :||||| :||||| :|||||
8730 TCTTGATCTCCTGCACGCTTCCTCCATCCACACGATGCGACGCAAC 8681
134 euserGIyGIyGIuGIuInthArgValThrleuAlaGyAlaLeuAlaVal 150
||||| :||||| :||||| :||||| :|||||
8680 TCTCGGCGGCGCAGACGAGCTGTGCTAGCGCGGCACTGTGATC 8631
151 AsnProSerValIleuLeuMetAspGIuProleuSerAsnleuGIuAla 167
||||| :||||| :||||| :||||| :|||||
8630 AACCCCAAGTCCTGTGATGAGCAACCGCTGTCCAACTTGACGCCAA 8581
167 sleuArgleuAspMetArgIuAlaIleArgGIuIleGIuInthIscIuValG 184

```

8580 ACTTCGCGGATTTGCGGCGGAAGTGGTGAATTCAGCAGCAGCATGA 8531
384 LyllethrthrvAltyrValThrhIsAspGlnGluLualamAlaIle 200
8530 ACATCCACCAACCGTTCATGTCACGATGATCAGGAAGCGCGTTCGATG 8481
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyLyr 217
8480 TCCGATCTGCTTGGCTCATGTATGCGGCGGTATTCAGCAGGCTCCACC 8431
217 gProLysGluLeuTyRHisLysProAlaAsnGluPheValAlaThrPheI 234
8430 GCGCGGGAAGTCTATACATCCCGCAACGCTTGTTGGCTCCTTGG 8381
234 leglyArGhrAsnIleIleProAlaAsnGluLysArGserAspGly 250
8380 TGGGAGCCATATATTCCTGACGCTGAC.....CGCACGGGTGGG 8340
251 AlaTyRlleValPheSerAspGlyTyRAlaLeuArgMetProAlaLeuAs 267
8339 CAACATTCATATCCAGC.....CTCAAGGTACGCTGCCGACGCGGA 8296
267 gAlnVal...GluGluGlnAlaIleHisValSerIleArgProGluGluP 283
8295 TGGCATTCACAGACGCGGACGATCGTGGCTGCGATCCGCCGAGGCGA 8246
283 heIle.....LysAspGluSerGlyAspIleGlu 292
8245 TTCTATGTCCTTCATAGGCAATGTGATGCATCGACATCGACGCC 8196
293 gLylThrIleArgAspSerValTyRleuGly.....LeuAsnThrAspTy 307
8195 GTACCATACGACGAGCTCACTGCTGCTGATGATGACCTGCCGCC 8146
307 rPheIleGluThrGlyPheAlaSerLys..... 316
8145 CGGCTGCTGCGGTGAGAAATGACGCGATCACCAACCATCGCGG 8096
317 .....IleGlnValSerGluGluSerThrPheGlu..... 326
8095 AGATCATCGCGCTTCAGCCACCAAGAAAGCATCTCTCTGTTGCG 8046
327 .....GluAspLeuGlnLysGlyAsnArgIle..... 335
8045 GCGGATACCAATTGTCGTCGACGCTGACCGGGAGCGCCTGTCATG 7996
336 .ArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
7995 ACGGATATGCGTCGACATGCAACGACGCGGCTTCGCCCTGCGAC 7950
seq_name: gb_AE001779
seq_documentation_block:
LOCUS AE001779 20990 bp DNA linear BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 91 of 136 of the complete genome.
ACCESSION AE001779 AE000512
VERSION AE001779.1 GI:4981777
KEYWORDS
SOURCE
ORGANISM Thermotoga maritima.
Bacteria; Thermotogales; Thermotoga.
REFERENCE
1 (bases 1 to 20990)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
Eisen,J.A., Fraser,C.M. et al.
TITLE Evidence for lateral gene transfer between Archaea and Bacteria
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316

```

```

REFERENCE
2 (bases 1 to 20990)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
location/Qualifiers
1..20990
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
complement(97..1203)
/gene="TM1224"
complement(97..1203)
/gene="TM1224"
/note="similar to PID:1208894 SP:044406 percent identity:
54.86; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, xyR-related"
/protein_id="AAD36299.1"
/db_xref="GI:4981778"
/translation="MPKSVRAENISRLIKRIMKSPVSVLEAEELGTLKTVGEIARI
PLEGIYVEEDSPKRGKRPKSLKISPNCAVYVIGIETVTRPIACLIDASMNILAH
AHLPSQSDRETELNVMTIRIDRAKDMMEKLGSLATVAAPIDIERKIITIDPRN
FPLSQIPLANLKEKRYGLEWVENDADMGAVGEKWKYRKDSFAMITLKGIGAGIIL
DGELYRGSENGAGEIGYRVNGENVLEEDVCENNVYLKHVLSMGFSSLEADSDG
VRKEYEFDIARFYSIGLNLILHFGISKIYVIGGFELGKNGENFLKIKIIEVETHLVK
HSVMSKSPVOEPIARNGAVALNLENTLERTTS"
complement(1217..2197)
/gene="TM1225"
complement(1217..2197)
/note="similar to GB:Pyro_h percent identity: 52.69;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAD36300.1"
/db_xref="GI:4981779"
/translation="MKVTEKIPNIPMEERPEGYTGPMWRYSKNPIIGRNPPGARV
ENSAVPPNGSEFVGRDKNTRPRLIFGSKGKGINNEIPEEIQWVDVNGEPQPS
YADPRVVKIEDTYITFTCDHGTIGVGMTKDKTFRLPNATVPNRNGVLPKPK
INGKVMNLRPSDNGHPFGDIFLSESDMTLHWGNHFRVLGRSSYNMWNELKIDAGPY
PIETSEGMLLIYHGVTLCNGYVYSFGAALDLDDPSKVLRSRYLLTTPREEYETVG
FVPNVFPCALCDADTGRVALIYGAADTHVALAFGYIDEIVDFKRRSM"
complement(2250..3938)
/gene="TM1226"
complement(2250..3938)
/note="similar to GB:Pyro_h percent identity: 57.09;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein, putative"
/protein_id="AAD36301.1"
/db_xref="GI:4981780"
/translation="MNMKRFLIVPVVFLFSSISQVLERNETWYGGSLSPSSNNN
PPTRNNAVPGTGLVETMFFYDELITGFPDLMALEKGMDELSKTRYVLRGTYSHDN
VPLTSEDVRFTELAKRYKGIHYSVWMLDIETPDNRVILFVKDRIHEMELLY
TLPLTKHIMEKETITLQSSNEVPLSGVVAHSWQNMKINERFENMGTVMYG
KPAKYVIVILVESLNNVALGMLKGEIDFSNMLPGVILKKVNLNLTXYDEPYHLS
SYVGEFLNARKYPLSLPEFRALAMSINADPIYORVYEGAVLADPLGFPYSVMK
XPKVEVKEHGEKYPDEPARKSLIDKLGFRVNGGSGFRFPDGKPKIKLITBCPGMTQY
MOAIVYIDQIAKVGINAEPIFPDSSKTYENMYGEPDIEMNANGTGISSTPWTYPT
IIPDPALESESYTGNGRYONPEVESLLEELNRPIPDNVEYELDGLGSELITLKL
FPIPLWGAFAFIQDNVYTWMPNEHNPVAMPCCWAMWQIGALKILFNLPKPAR"
complement(3935..5944)
/gene="TM1227"

```

gene

CDS
complement(3935..5944)
/gene="TM1227"
/note="similar to PID:1679597 percent identity: 54.44;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endo-1,4-beta-mannosidase"
/protein_id="AAD36302.1"
/db_xref="GI:4981781"
/translation="MRFMEILLIVELSELVFLASDEFVAVKENGKALNCKERRRIGTSN
NYMYKSNRMIDSVLESARMDGKIVLRIGWGLDSEYCRDNTYMHPRPGVGEK
ISNAOSGPERLDYVAKAKELCIKLYIVVNMDDGGMNQVYRRPGGHNDHDFDE
KIKREKTKVSVFLVNHNTYGVPTREEPTIMAMELANEPCETDSSGNTLYEWMKEM
SSYIKSLDPNHLVAVNDEGEFSNIEGFKPYGGAEWAMVMSGWKMLLSTLETVDG
FTHLYPSHSDPENVAGQWAKMIEDHIIKAKIIGKPVYLEEGIPKSAVNRATAYR
LNNDLVYDLGGGAMFWMLAGIGESGDRDERGYPIYDGFRIYNDSPALRIEXAK
LNTGEDIREDTCSFLPKDGMIEKTKTVRAVGVPSYNTPEKLSVYKVDIAEENIE
HIGYGVGFDLDTTRIPDGEHEMFLGHCQCKVYKSKAKYVNRARYLAEEVSS
PREVKMANSSTWQAEFGSPDIEKNGEVCNGALQLNKLPGKSDMEVYARKFERLS
ECEILEYDIYIPNVEGLKRLRPVALNPGWVIGLDMNNAVESLEITVFGKEYRR
FHVRIEFDRTAGVKELHIGVVDHLRYDGPFIIDVNLVYKRTGGM"
complement(6103..6903)
/gene="TM1228"
complement(6103..6903)
/note="similar to GB:M81878 SP:E26833 PID:144859 percent
identity: 55.56; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, RpiR family"
/protein_id="AAD36303.1"
/db_xref="GI:4981782"
/translation="MKIRDKILNVYTFQSPAEKRYADYLERPDVIRHSITEFAKIV
GVSEETIHRMIKKLDEGEYQAKIALARELSGLEETIERDFIDEIDIRLKLTD
MKNVKAVMIISARHILFEFGVLSGVSEYSAIKFSLGFRHTSPNDHYVIAVAV
LTGEDLVISHTGNIRDTVKSTQVAKDGAKTATITTKOSELKAVHVLVQSPVK
IDTYEFLRENIEIAVVDYLERFQRIYREKKHFNLEGVFKPRF"
complement(6887..8335)
/gene="TM1229"
complement(6887..8335)
/note="similar to GB:Pyro_h percent identity: 69.66;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36304.1"
/db_xref="GI:4981783"
/translation="WESASGSRSOKARGAFRVNSLQIRYKTSRPEIRROSRMLQR
ERRPKGGRDEFGKRGNFSPLRVSLQDGRQLRSKELQFRRRHRHEDGKIT
MKVVGIPSYNAEETISHVARTAAQIYDFEGDAIVSDSGADGTERHMDGTF
GLPKESFYEGGLPKGSAMRAIMEFALAKODAAVVELDLDSVPMWNERLAGVELK
GEADYVTPYLRHRFDGTTNNVCPMPATVGLGKRVROIGDFFGKRLITLYGKP
KRIWMDVARRGIDIMTMTTAINESGRVQAOLGKRVHVKRQKHKMCFQVGTG
FELVITYENWKEIKKIEDVPIYIGETPOEVPKSIDIGNLKLARETLEEVEYIDRG
ILSEVKESGTLSSWVDLILRSAAVOYRTRTKVYENLLPPEYFARTAFABEVSLS
DEAEARYVEQDLVLEKHSLEEMKEDKR"
complement(8026..9309)
/gene="TM1230"
complement(8026..9309)
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36305.1"
/db_xref="GI:4981784"
/translation="MKIALIHRGGLMDVSLSEMKKKVLTIRMGHEVHIVAKNKEG
VDLTLEIGFENPDERVNRNFGGKIDFLSEKEDFLKEKEEELFIHLNALDYD
LIVPNINWISGLPSLGLASLEKRFVAAHNDHDFMERKHILIPKRRPEETILKRP
DLPNKHYVINTIAQBELKRRRINDSVYVNPVMDSPITSEEMTHRVREELQIAPGT

IVALQATRIDRRKTELSDVSLKELTITSKKADLYNGERYSEVILLSEGTDE
EYKLEKEVASSGVSLVISEVKNRNSLFEKLLNADFPVYPSILBGMQMLEAI
AAKKPVVLFEEYVFSKSDIKPACIKVYSLDRCFRNGVYKVDERILKRAVEISRLF
DPSLRTEVHEHFEVGRHFSLERLEDLSREVLD"
complement(9306..11804)
/gene="TM1231"
complement(9306..11804)
/note="similar to SP:p21139 PID:205299 percent identity:
49.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-mannosidase-related protein"
/protein_id="AAD36306.1"
/db_xref="GI:4981785"
/translation="MKVYVYVNHMDREMTSSEYTSKMLKEVFRVKEVYOKNPEV
YLDQOTAPAEVLVYHDLDEDLRELVRSGSLLAGPYIYIDMFIPEGASTLKNEI
GKDPNRRGRBNACMLDSRCHTISOEOLHPICTEVEFLMRGISFENDGISOEPW
KSDGTAAQGVTVLGGYRNLVNLAKETODIAERLKHVEYKLAKEFRSGEILLDGYD
DISPDPADYLNVEIVSPPEPERPENAPVTLTSHDSCGACADVPVHEKMERVRY
SOLVERKLKMDLSVLNGEELSDFWMDYLTSLAHNDSGVCADVPVHEKMERVRY
YFLVKSVESSRLNHLVERGLGKGTALSLSPSYNHHCDERTFLKTNGAGLEKT
EVADEKKVEYKELTMENERYRAYFEGDGTLYLNRRAGVYKLEKEDGAYTSSVEETD
SISLEKMPVLOTKSMVYELERTIRSSOCLITEMERVYIDETPLVKKRIALIKIKGN
YLSWVYETGSEITRAKMPDYVYKREKDSYLLPTPEBELKGLILAAKEGEVERP
FOGFVSVSESSKTIILARGLKRYMEEGKAVHTLSRVENYLAKEVKKRVGDAGPMY
VPGACGEVFTVDALFMEELPVPRSEEEFRWLLDNDPDLADVADGTEKKTFFS
SLPMAGVGDYRLVYVNYLEIEICLRLPLOGKSIESIASEOCVSRNLINPEEP
VYVSEKNPEGEVIEFLKELTSLSEDEITNLBECKDPRSKHRLFSLLRKKELISLOY
LIDVKTARELNELMKRRTYVYVLEFSEERE"
complement(11801..12868)
/gene="TM1232"
complement(11801..12868)
/note="similar to GB:Pyro_h percent identity: 78.03;
alignment_scores:
quality: 716.50 length: 363
Ratio: 2.559 Gaps: 6
Percent Similarity: 77.135 Percent Identity: 40.220
alignment block:
US-09-769-787-162 x AE001779/rev ..
Align seg 1/1 to reverse of: AE001779 from: 1 to: 20990
1 MetserGluileuysileileasAlaLysylsletyrHIsapValPr 17
12868 TTGGCCGACGTGAAATAGACGGAGTGAACAATATTTCGGGAATGCGAG 12819
17 ovalIleGluAsnleuAsnIlethrIleProLysGlySerLeuPheThrL 34
12818 GGCTCTGATGAGATCGACCTTGCCTGGAACAAGAGAGAGTTCCTGCTG 12769
34 euLeuGlyAlaSerGlyLysThrIleLeuAlaArgMetIleAla 50
12768 TTCTCGGTCCATCTGCTGTGGAAAGCAACCTTCGAGGTGTATAGCG 12719
51 GlyPheAsnSerIleGluGlyGluPheThyrPheAspThrLysI 67
12718 GGTCCTGACACAGTGACGGGGGAAACATCTTTTCAACATCGCGAGCT 12669
67 eAsnAsnMetClnuProSerLysArgAsnIleGlyMetValPheGlnAsnT 84
12668 GACCAACCTTCCTCCAAAGACAGACATTCGATGCTCTCCACAGCT 12619
84 yralIlePheProHisLeuThyValArgAspAsnValAlaPheGlyLeu 100
12618 ACGGGGTGTGGCCCATGATGATGACGACAAACATCGGCTATCCCTGTG 12569
101 MetClnLysLysValProLysGluGluIleGlnIleThrAsnLysT 117
12568 AACTGAAAGAGTCCGAAAGAGAGATGAAAGAGGGTAAAGTGGC 12519

```

117 rleugluuemetGlnllealaglntyrAlaasparGlysProaspLysL 134
    :::::::::::::::::::: ||||| ||| ::::
12518 AGGGATCTCCTTCACATCTCCGAACCTCTCGACAGTACCCGACAGC 12469
    :::::::::::::::::::: ||||| ::::
134 euSerGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
    :::::::::::::::::::: ||||| ::::
12468 TCTCCGAGAGTACAGACAGAGGTTGCCGTCGCCAGCGATCGTTAC 12419
    :::::::::::::::::::: ||||| ::::
151 AspProSerValleuemetaspGluProLeuSerAsnLeuGluAlaVal 167
    :::::::::::::::::::: ||||| ::::
12418 GAACCCGAAGTACTCTGTATGACGAGCCCTTTCCAACTGTGAGCCACT 12369
    :::::::::::::::::::: ||||| ::::
167 sleuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
    :::::::::::::::::::: ||||| ::::
12368 TTTGAGAGGTGAGATGAGAGAGCAGCTGAAAAAGCTTCAGAGAAAGATTG 12319
    :::::::::::::::::::: ||||| ::::
184 lylleThrThrValTyrValThrHisAspGlnGluGluAlaIle 200
    :::::::::::::::::::: ||||| ::::
12318 GAGTCACACCATCTATGTGACTCATGACACAGACAGCCATGACGATG 12269
    :::::::::::::::::::: ||||| ::::
201 SerAspGlnlleAlaValMetLysAspGlyValIleGlnGlnleGlyAr 217
    :::::::::::::::::::: ||||| ::::
12268 GGGACAGAGATCGGGGTGATGAACGAGGAAAGCTTCAGAGGTGGAAC 12219
    :::::::::::::::::::: ||||| ::::
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234
    :::::::::::::::::::: ||||| ::::
12218 GCCCTGTGAGATCTATCACCATCCGGTGAACATCTTCGTCGCTTTTG 12169
    :::::::::::::::::::: ||||| ::::
234 leGlyArgThrAsnIleleProAlaAsnLeuGluLysArgSerAspGly 250
    :::::::::::::::::::: ||||| ::::
12168 TGGGATCACCCGAGATGAACCTTCTTAGAGATGAGATGAGATCGAAGA 12119
    :::::::::::::::::::: ||||| ::::
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
    :::::::::::::::::::: ||||| ::::
12118 AATAGCGTTGTCTCCAAAACGGT...GAGATTAAGATTCCGCCGAAAGAC 12072
    :::::::::::::::::::: ||||| ::::
267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluGluPheI 284
    :::::::::::::::::::: ||||| ::::
12071 CGATCCAGGTGCAAAAAGCTGATCTTGGTATTCCAGCCGAGAACGTCT 12022
    :::::::::::::::::::: ||||| ::::
284 leLysAspGluSerGlyAsp.....IleGluGlyThrIleArgAsp 297
    :::::::::::::::::::: ||||| ::::
12021 AATCGAGGAAAGCCGAAACACTTGAAGTGTGAAGA..... 11984
    :::::::::::::::::::: ||||| ::::
298 SerValTyrLeuGly.....LeuAsnThrAspTyrPheIleGluThrG1 312
    :::::::::::::::::::: ||||| ::::
11983 GAGGTTTATTTCGCCGAGAAACTCATGTGATACGATTCCTCATCTCAA 11934
    :::::::::::::::::::: ||||| ::::
312 yPheAlaSerLysIleGlnValSerGluGluSerThrPheGluGluAspL 329
    :::::::::::::::::::: ||||| ::::
11933 CGTTGGAAGTAG..AAGATCTCCCAAAAATTCAGAGGGAGTGTGATT 11887
    :::::::::::::::::::: ||||| ::::
329 euGlnLysGlyAsnArgIleArgLeuArgIleAsnThrGlnLysLeuAsn 345
    :::::::::::::::::::: ||||| ::::
11886 TCAGAGCGGTGAGAGATCACGTTTTCCTGAGCGTTGAAAAAGATCCAT 11837
    :::::::::::::::::::: ||||| ::::
346 llePheSerAla..AspGlySerGlnAsnLeuIleLys 357
    :::::::::::::::::::: ||||| ::::
11836 CTCCTTACCCTGAGACGGAGAGAGATCTCATGAA 11799
    :::::::::::::::::::: ||||| ::::
seq_name: gb_ba:AP003189

```

```

seq_documentation_block: 298050 bp DNA linear BCT 17-JAN-2002
LOCUS AP003189
DEFINITION Clostridium perfringens DNA, complete genome, section 5/10.
ACCESSION AP003189 BA000016
VERSION AP003189.2 GI:18146727
KEYWORDS
SOURCE Clostridium perfringens (strain:13) DNA.
ORGANISM Clostridium perfringens
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.

```

```

REFERENCE 1 (sites)
AUTHORS Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
        Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
TITLE Complete genome sequence of Clostridium perfringens, an anaerobic
        flesh-eater
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
PUBMED 11792842
REFERENCE 2 (bases 1 to 298050)
AUTHORS Shimizu,T.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
        Sciences, University of Tsukuba, Department of Microbiology; 1-1-1
        Tennodai, Tsukuba, Ibaraki 305-8575, Japan
        (E-mail: tshimizuend.tsukuba.ac.jp, Tel:81-298-53-3354,
        Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18144662.
FEATURES
        source
            location/Qualifiers
                1..298050
                /organism="Clostridium perfringens"
                /strain="13"
                /db_xref="taxon:1502"
                /note="anaerobic pathogen for gas gangrene"
                /complement(34..1036)
                /gene="CPE1002"
                /complement(34..1056)
                /gene="CPE1002"
                /note="340 aa, no significant homology."
                /codon_start=1
                /transl_table=1
                /product="hypothetical protein"
                /protein_id="BAB80708.1"
                /db_xref="GI:18144663"
                /translation="MSNKKIIFKNNNSDKVLIINSLVLDGSLAKELSKKIDIOKRD
                IIKMNEBGFDDSDGYFKDKDEESISIDIPVEYSKSLPFPREDVQSIIDSI
                NEKLNIDIEFTNLLPEIKTSEKVSATKEDNDNTKTSQSMLEIPVRSIARRK
                KKKKTLIRMIISYFNKNSFKFLIKKENKSSKTECDNENKKEVNIKSTNIIIE
                VAIEKLIVNDKIETTSGLNSESRLPEIKIPKFKGRRRIPKVDCLDEINFLNSD
                YSLDEKKDFDLKLSDFIKNINSDSLESNRDLSLEBRVONLEKKIINLEEFLEIF
                QAMSNK"
                1361..1540
                /gene="CPE1003"
                1361..1540
                /gene="CPE1003"
                /note="59 aa, similar to gnu:AP001515_87 BH2353 gene
                product from Bacillus halodurans (79 aa); 46.7% identity
                in 45 aa overlap"
                /codon_start=1
                /transl_table=1
                /product="hypothetical protein"
                /protein_id="BAB80709.1"
                /db_xref="GI:18144664"
                /translation="MNIDELTKRINELKKHKEGLSEDEHKEDELRKEYINRFKSN
                LRQLKGLPEPKNNKN"
                1552..2475
                /gene="hprK"
                1552..2475
                /note="307 aa, similar to gnu:AP001519_105 HPr (Ser/Thr)
                protein kinase/phosphatase from Bacillus halodurans (310
                aa); 45% identity in 282 aa overlap
                CPE1004"
                /codon_start=1
                /transl_table=1
                /product="probable HPr Ser/Thr protein kinase/phosphatase"
                /protein_id="BAB80710.1"
                /db_xref="GI:18144665"
                /translation="MGVTIEKLIDFSLEVIQIGENVPINVDNRPGLQLAGFYNY
                FAPRIQVIGKAEKSPLEDMSPDLRRKRINKFPDLSCLITGLIEHEELAKARK
                RNWILSDSWYTFPISKITMYLSDKMAPETRLRGVLDVYGIOMLITGSGIGKSTP
                ALLEIKRHLRVTDVAVDIKETDIDLGRSEITFGMLKVRGKGIITDYSALIGSSTL
                NSKQIKITIEHMKDGDGVDRLGVNDVYODILGVKKKLKLVPIRPGRNLAIVITEAA
                ANYRQRMSDISPDIIEKRLSEMESESKI"
                2520..3449
gene

```


KEYWORDS	AE001791.1	GI:4981929
SOURCE	Thermotoga maritima.	
ORGANISM	Thermotoga maritima	
REFERENCE	Bacteria: Thermotogales; Thermotoga.	
AUTHORS	1 (bases 1 to 18251)	
TITLE	Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uitterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., Eisen,J.A., Fraser,C.M. et al.	
JOURNAL	Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima	
MEDLINE	Nature 399 (6734), 323-329 (1999)	
REFERENCE	2 (bases 1 to 18251)	
AUTHORS	Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uitterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	1..18251	
gene	/organism="Thermotoga maritima"	
gene	/db_xref="taxon:2336"	
gene	92..832	
gene	/gene="TM1368"	
gene	92..832	
gene	/gene="TM1368"	
gene	/note="similar to GB:PYro_h percent identity: 70.83; identified by sequence similarity; putative"	
gene	/codon_start=1	
gene	/transl_table=1	
gene	/product="ABC transporter, ATP-binding protein"	
gene	/protein_id="ABD6439.1"	
gene	/db_xref="GI:4981930"	
gene	/translation="MLKIVLHAKLRDEDEKILKVNLEIEKEBHVHLMGPNQSGKSTLAVVINGNPRIVYIVTEGDIVPEGNSIKDLPENRKLGIIMTFQNPYEVEGVKLSQFLITARRIHGEDKNYLERLEKELETAEKIGKDLFLELYLVNGFSGEKKRSEIIQSFLRPLLIIDEISGIDVALRLINLRLINLRLINEGVTLITLITRYKRLDHLKRIDKHVYVDRIVTSGGPELADVEDIEEKYSLEGR"	
gene	829..2223	
gene	/gene="TM1369"	
gene	829..2223	
gene	/gene="TM1369"	
gene	/note="similar to GB:AL009126 percent identity: 74.45; identified by sequence similarity; putative"	
gene	/codon_start=1	
gene	/transl_table=1	
gene	/product="conserved hypothetical protein"	
gene	/protein_id="ABD6440.1"	
gene	/db_xref="GI:4981931"	
gene	/translation="MMERLITDSDRFNFVPRKVTAYKAPGCLDEKLMETISRAKDEPEWMKLRLESLKVNEMHNPPFGVDISGLDGKIVSYTKPPAKKSTSEWEPPEYKEAFDKGIIPAEKRYLAGVAQDSEIVYQNMKELEKMGVITLDMESAAREPDDLVKKEFMKVLPIIDHKFFAALHGAIVSGGTFVLVPAGVKIPMDQAVFLMSAPMGQPEHTIIAEESEVETVIGSGCAPRYNIINLHAGMEIVYKGAQAKKYLITONMSKNTVNLTKRSIVDEGSEWTWVSGISGQKTMILYPMPTILKGAARESSITVYAGGQHMDSRVVHLAIPYSSIVASKSTISGGMNAPRGCIKTKTRAVVSKASVECAALMDNRKSDNPVPIIEVEPDVGVGHEARIKRGIDQIYITLMSRGLSEDAKAMIVYKGVPEVVKELPREYAVELNKLEIEIEKSG"	
gene	2234..3340	
gene	/gene="TM1370"	
gene	2234..3340	
gene	/gene="TM1370"	
gene	/note="similar to percent identity: 0.00; identified by sequence similarity; putative"	
gene	/codon_start=1	

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36441.1"
/db_xref="GI:4981932"
/transl_table="MEKTLVIRHDEGKAVVETPOLIFSNPDYEVLEKALPNTGCP
KMKRKKTOEYKEMGFPRKRCNDGMSLPESFRIDIDVDISLEKIDEGARKE
VLLGDTFFDFVYDGHKMMRFDDGILENVLVVEKEAETIKLSSTDRNTVMRIL
VKKNSLKFENLDFGNSFDNVFTVLEEGARLVJREFKAFGRKGTGHLVKKSSK
ADMSFEFYONGSGITDILLYLMRFEGEDAGKLGAGVYDAGKIVFRGILDYKRGSKN
IVAEEHTLVISPDARMDAIPSLWDENDVYASHSSASSSLDEDELFIYLMRGVDEI
ESKRLIVGVFELLDELDESKEVENVNIIV"
3318..4559
/gene="TM1371"
3318..4559
/gene="TM1371"
/notes="similar to PID:1001734 PID:1001708 percent
identity: 61.37; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="aminotransferase, class V"
/protein_id="AAD36460.1"
/db_xref="GI:4981951"
/transl_table="MLSTGLYEDPPTLMKINGKRIYIDSAATTLKROVYKLEEF
YNSYANVRAVHTLAVRATTEFEETREFEAFELGASPEILFTSGTMAINLAVSL
LNSGILKEDDLVLSLEHNAFVPLRLSKPHGRIDFIRPSGRFGLEMDLKHR
DKNPVVAITGSLNVTGORIPVEELRNVPNSIILLDGAQLIPHPVPKKEIGVPLA
FLHKMLGPTGVGLYKREVELEOMEPFLYGEEMIDRVSLIEDVTNENLPKFEAGTPN
IADIVASREALRALYKETGDSVYHEKITEOLTOMALKEMKIGGVETIYGLPDEROHIVS
FNKGIHHDVAHITDDEGCIIVRSRHHCAOPLMSILKEQSOIDFPNSTCRASFTIYN
TEEDVKILVEGKKVKEWFSR"
4544..4972
/gene="TM1372"
4544..4972
/gene="TM1372"
/notes="similar to GB:U00013 PID:466875 PID:2398702 percent
identity: 35.07; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="nitrogen fixation protein NifU-related protein"
/protein_id="AAD36442.1"
/db_xref="GI:4981933"
/transl_table="WVFKMYSEAILDYANSKRRKGLDADTVIEGKNISGDEITL
YLVKVDGVKDAKFEKMGCVISQASASLIERIIGERVEEIRSLIEAEKMSRGENFD
EGKLNVTLMSPKIKNVPARVKCFILAMTKLTKALKISR"
complement(4950..5936)
/gene="TM1373"
complement(4950..5936)
/gene="TM1373"
/notes="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36443.1"
/db_xref="GI:4981934"
/transl_table="MRKLSIVFLIVASLVFMSGHEGLTYIVKDIVDEWVITTPY
NYOENRVNLSKLVLEDVCGDDTISAMGALIPRPSPEDGVPAMOLITVYSSPDW
GMDGSLDISPYOFLIGNSQVNHMKYRITGLFEFFADKSFTYFHMAKAPRKGRYV
GTRFLARALHYTELSQPHNTPGVEGELFQALFDRSITMKMIMNAHYVDDVLATLY
SGDEETIKTEEATVPFLKSEHNLVQEVKRLASTYPRVHREIKVFEKLEKERSDL
ESFVIVSEEFREADQRGDVLVKRYTLELICKVSSYLKGYLIDFLKAVKSP"
5984..6589
/gene="TM1374"
5984..6589
/gene="TM1374"
/notes="similar to GB:AE000657 percent identity: 62.38;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase"
/protein_id="AAD36444.1"

```

```

/db_xref="GI:4981935"
/transl_table="MKLYIRHGETINNEKGLMOGYVDYDPLNREGEQARKLANSLKR
VDAYSPLKRSLETAEEIARFEKEIIVDEEDICEISLNNGLVEEALREYVEFK
KMSDPNFGMEGLSEMRVONRVVYAIKIRYIOEKLNSSENVYIVSHLSIRAFICWI
LCGLPLVIRNRKIDNASLVYIEESKPRVLVLLNDTCHLKE"
6627..6700
/gene="tRNA-Cys-1"
/notes="codon recognized: UGC"
/product="tRNA-Cys"
6627..6700
/gene="tRNA-Cys"
6756..7766
/gene="TM1375"
6756..7766
/gene="TM1375"
/notes="similar to GB:M64519 SP:P23861 PID:147329 GB:U00096
PID:1651550 percent identity: 60.95; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="spermidine/putrescine ABC transporter,
periplasmic spermidine/putrescine-binding protein"
/protein_id="AAD36445.1"
/db_xref="GI:4981936"
/transl_table="MKRVFLVLVLTALPIPAKATLYIYMADYIPEDYIRAFEEKY
DCRVYDVTYASNEEYAKFKAGCGGYDLIPSGDYVIMKREGLQKLDLSKIPNFK
YLDKDIILAKTYDDPHNERSVYPMGMSYIVNKKVYKESWSIFDRDLRGRTLL
DDMRVLAALAKLYGLSVNTKMPKLEBAKOVYLMKKINAKFDASSYADGIVSGEYV
VYHGAEEVQYVIRPGEKRYDFEIPKREGTLMIDNVIIPGAKNVDLAYKFINILE
PENNAKIDYIGLSPNVEAKRYMOTEPYITIEDLKNEFTIEDGELALEYKNTMLEI
IM"
7781..8887
/gene="TM1376"
7781..8887
/gene="TM1376"
/notes="similar to GB:M64519 SP:P23858 PID:147326 GB:U00096
PID:1651555 percent identity: 67.38; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="spermidine/putrescine ABC transporter,
ATP-binding protein"
/protein_id="AAD36446.1"
/db_xref="GI:4981937"
/transl_table="MIGGEVSIKYNKSFEDDFOVULKNSLDIKKEFPSSILGSGCGK
TTLAKVINGFEVSGDVLLDQKSLTNLPKRNVPNIIFONVAFPLHLYVENIAFL
KIKKISEINQVNWELSLIMEBNAOKMSOISGGOKNOVALARALANPRVLLLD
EPKSLDKKLNQELLVELDNLHDKRGITFTYTHQDAISVSDBVALMNGEITVOVG
TPYEYVESPVNVFAATFIGETNLMAEVEVEDEYVVESFGIGFCYRDKKEAKGD
RLILTRBEKIRISRKOPRSETEFNVFVGVDDELYMGHQTKYFRVLDGXYIMRYKQ
HARYILDEPIIKMEDEVETWNPDSEFIVEVLEE"
alignment_scores:
Quality: 701.00 Length: 327
Ratio: 2.804 Gaps: 4
Percent Similarity: 76.453 Percent Identity: 42.202
alignment block:
US-09-769-787-162 x AE001791 ..
Align seg 1/1 to: AE001791 from: 1 to: 18251
3 GtU1leUys1leUleasna1aUys1leUyn1sAspVal1ProVal11 19
|||||:||||| |||:|||||:|||||: |||
7793 GAAGTTTCATTAAGAACGTCACGAATTTTTCGATGATTTCCAGGTTT 7842
19 eG1uaSn1euaSn1leThr1lePro1ygs1ySer1eupheTh1leu1eug 36
|||||:|||||: ||| |||:|||||:|||||: |||
7843 GAAGAACGTTTCCCTGGATATTAAGAAAGAGAGATTTCATCTCAGTGC 7892
36 1y1a1eSer1yCys1y1ySer1Th1leu1euar1yMet1le1a1y1yPhe 52
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7893 GCCCATCGGGTGTGGAAAAACAACGCTTCCTGCTGTGATGAGCGGCTTC 7942

```



```

53 AsuSterileGluGluGluPhePheAspThrLysIleAsnAs 69
   :::::::::::::::::::::
7943 GAGGGGTGGAAAGTGGACATGTTCTCCGATGGAAAGTCGACTGAA 7992
   :::::::::::::::::::::
69 nMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaI 86
   :::::::::::::::::::::
7993 TTGGCTCCGAAACAAAGCCCGGTGAACATCATATTTCAGAACTACGCTC 8042
   :::::::::::::::::::::
86 LePheProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGln 102
   :::::::::::::::::::::
8043 TGTTCCTCATCTCACAGTTTGGAAACATCGCTTCCCTGGAACATC 8092
   :::::::::::::::::::::
103 LysLysValProLysGluGluLeuIleGlnInThrAsnTyrLeuGln 119
   :::::::::::::::::::::
8093 AAAAAACTCCGAGACGAAATATCATGAGATGACGAACTGCTTTC 8142
   :::::::::::::::::::::
119 uLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerG 136
   :::::::::::::::::::::
8143 GCTCATAAAGATGAAAGAACACGCTCAGAAATGCTCCCAACTGCTG 8192
   :::::::::::::::::::::
136 LysGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaAsnPro 152
   :::::::::::::::::::::
8193 GTGGTCAGAGAGAGAGTGGGATGAGGAGGCTCCGCAACGAGCG 8242
   :::::::::::::::::::::
153 SerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArg 169
   :::::::::::::::::::::
8243 AGGGTCTCTCTCTCGATGACCGCTGACGCGCTCGACGCAAACTGAG 8292
   :::::::::::::::::::::
169 gluAspMetArgGlnAlaIleArgGlnIleGlnIleGlnValGlyIleT 186
   :::::::::::::::::::::
8293 ACAGCAACTCTTGTGACGCTGACCAACTTCATGACAGTGGATGATCA 8342
   :::::::::::::::::::::
186 hThrValLysValThrHisAspGlnGluAlaMetAlaIleSerAsp 202
   :::::::::::::::::::::
8343 CCTTCATTACGTAAACACAGATCAGCTGAACCAATCAGTGTTCAGAC 8392
   :::::::::::::::::::::
203 GlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 219
   :::::::::::::::::::::
8393 AGAGTGGCCCTCATGACGAGAGAAATGCTTCAGGTTGGAACTCCCTA 8442
   :::::::::::::::::::::
219 sGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyA 236
   :::::::::::::::::::::
8443 CGAGTGGATGAGAGCCCGGTCAACGCTTTCGACGACCTTCATCGAG 8492
   :::::::::::::::::::::
236 rGThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyr 252
   :::::::::::::::::::::
8493 AAACGAACCTCATGAAAGCCGAGGTTGTTGAAGTGAAGTGAATATATAC 8542
   :::::::::::::::::::::
253 IleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAspGlnVa 269
   :::::::::::::::::::::
8543 GTTCGTGAAGAACCGGGAATCGGCCAGTTCAGGTTTACAGAGACAAGA 8592
   :::::::::::::::::::::
269 lGluGluGlnAla.....IleHisValSerIleArgProGlu..... 281
   :::::::::::::::::::::
8593 ACCAAGAAGAGGTGATGCTCATCACTTACAGCCGGAAGATCA 8642
   :::::::::::::::::::::
282 .....GluPheIleLysAspGluSerLysPhe.....Glu 292
   :::::::::::::::::::::
8643 GAATCTCAAGAAACAGTTCGTTCCGAGAAACCTTCATGATATTTCAC 8692
   :::::::::::::::::::::
293 GlyThrIleArgAspSerValTyrLeuGlyLeuAsnThrAspTyrPheI 309
   :::::::::::::::::::::
8693 GAGGTGTGGATGAGCAATATATCATGGGCGACCAACGACAGTACTTCGT 8742
   :::::::::::::::::::::
309 e.....GluThrGlyPheAlaSerLysIle 317
   :::::::::::::::::::::
8743 TCGACTCGACGAGGCTACATATGAAGTT 8773
   :::::::::::::::::::::
seq_name: gb_ba:AF307052
seq_documentation_block:
LOCUS AF307052 18001 bp DNA linear BCT 12-FEB-2001

```

```

DEFINITION
Pyrococcus furiosus transposase, sugar kinase, trehalose/maltose
binding protein (malE), trehalose/maltose transport inner membrane
protein (malP), trehalose/maltose transport inner membrane protein
(malG), putative trehalose synthase, trehalose/maltose transport
ATP-hydrolyzing protein, putative sulfate transport system permease
protein, putative sulfate transport integral membrane protein,
putative sugar-binding transport ATP-binding protein, putative
solute binding lipoprotein, and transposase genes, complete cds;
and unknown genes.
ACCESSION
AF307052
VERSION
AF307052.1 GI:12018039
SOURCE
Pyrococcus furiosus.
ORGANISM
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1 (bases 1 to 18001)
Diruggiero,J., Dunn,D., Maeder,D.L., Holley-Shanks,R., Chataud,J.,
Horlacher,R., Robb,F.T., Boos,W. and Weiss,R.B.
Evidence of recent lateral gene transfer among hyperthermophilic
archaea
Mol. Microbiol. 38 (4), 684-693 (2000)
JOURNAL
2056786
MEDLINE
11115105
PUBMED
2 (bases 1 to 18001)
REFERENCE
Dunn,D.M.
Direct Submission
Submitted (20-SEP-2000) Human Genetics, University of Utah, Rm 308,
BPRB, 20 S. 2030 E., Salt Lake City, UT 84112, USA
FEATURES
source
location/Qualifiers
1..18001
/organism="Pyrococcus furiosus"
/db_xref="taxon:2261"
60..840
/organism="Pyrococcus furiosus"
/insertion_seq="IS-like element"
/db_xref="taxon:2261"
84..833
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="AAG45369.1"
/db_xref="GI:12018040"
/translation="MSGISVWYFRKRGRGNKMSFTIIYVVSALKPPRRNKIPPEKTI
RGVELYLRGLSYROTARILKLSHTYWEAVGKLAEAYKKRLVYKKNRIANDERY
IKIKGKRITLMAAIDVESKEVLAIVITVRKMWARQDTILVLSGSGQPFVLDKRS
WYKSAFSLRGLYLVTFEGPRNSVERWFRLKEKTKRPWNFRGKDWRRVHRFVFLFA
FWYNFVRHSFGDPDPDYTEMLOEVMPOLS"
903..1220
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45370.1"
/db_xref="GI:12018041"
/translation="MKVAVVYLFLQSFILYYSARGSERVMLIFGLNFIAMGI
VKDRRAKRTIAKGYDFPALLMNGALIFQSLNAIDLAIIHDLIGLFGKKEPT
EES"
complement(1204..2142)
/codon_start=1
/transl_table=11
/product="sugar kinase"
/protein_id="AAG45371.1"
/db_xref="GI:12018042"
/translation="MIYATGELILDFLAKEGKLDVREEFKHPGAPANVVGRLRL
GAKSALISKVDDPGEFLIEBLKERVETRYIIKDNKRGIVFVOLIAGKPEFITY
DGVAVFNLREKEIOWDEMRDALELHFGSVLARPSSSTVEVLRVAVGKVPISYDVN
IRLDLNRGERKEMLDIEALKLADIYIGDELEYLNKNGIALEDFALVATLRGA
EGSPTIIKRDIVDVPSPKVPVDTTGADAPMAALIASLFTMGKLDLIEFSKEELKEL
GSFANLVAALSTTKRGANVSFSLSEVLKRRFSLP"
2376..3728
/gene="male"
2376..3728
CDS
gene
CDS

```

```

/gene="male"
/codon_start=1
/transl_table=11
/product="trehalose/maltose binding protein"
/protein_id="AAG45372.1"
/db_xref="GI:12018043"
/translation="MKVKKVLGLFVLGVGLAVVASGCIIGQOISTVISTPPIESLQ
GKIVEFVAGAPNEIEYWKVIAEFKKYGVYELKROATDEQRRLDVLNLRKSS
DPDVLMDVAMGQFIASGMLPELDYVORQMDLVSPFQSVINLADKQGLKALPY
YIDAGLLYRKDLLEKYSGSPETMOELVMAOKIOSGERETNPMEGFWMOGKQYE
GLVCDVEYVYVNGSGISGFKDGKWPVLNKPENVALQPMVDLHKYKISPNKYTE
MTEPRRLMEQOQNAFERNWPAMGLNADSPYKGVAPLPHFPHKSNATLGG
WHIGISKDNKALAMEVKEVSVQKGFAMNLGMPGRVDVDDPAVYKSPSLK
ELRAVENAVPRIVYVYQPSLSEIIQKYNVNSLAKGISQPEALDKAKQKAEELVKQYS
"
3770..4663
/gene="malF"
3770..4663
/gene="malF"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane
protein"
/protein_id="AAG45373.1"
/db_xref="GI:12018044"
/translation="MDNLLTSKLYREAKLGYLMILPLTTLVVLFIILVMTGFWISL
HSDVETPEKPEVGLRNLVRLVSAREFVSTFVTSFVSLETILGLSPALINE
RLKRGVRLAIVLIPWAPVTIISARTMELMNYQALFNMWISLIGSVNMLAPGIS
AFPAIYADVWKTTPETMTLLAGLQAIPODLYEALIDGASMEFEKRSITLPLKPV
LVALILRTIDALRPFDITTYLTGGSGGATTSISLARNTNLGDYIGSAISLTLF
VLVLSFTIYLVGRRREGLK"
4660..5496
/gene="malG"
4660..5496
/gene="malG"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane
protein"
/protein_id="AAG45374.1"
/db_xref="GI:12018045"
/translation="MREYVLRKRLILITGLAILMAITLFPFIMIVSFPEDPFLGSP
LVEKSTLENYRVLSDPYLHPFAYLAKNSIIIASLTTLTTSISSIAAASVRIEKG
RLILPIEVLGSMFQDISLVGLFKFIEKLGWNTYQALFPYVAWTLPLSLMITLSTY
FSQLPDLDEAAMIDGASRIKTLTIIPLSAPALSTALVFIAAFNEFMALLFTT
DHRARTVPGIALFGVHGEIPIWGSYMAVSISTPIVLIMALLFQKIVYSGLTAGALK
GE"
5501..6739
/codon_start=1
/transl_table=11
/product="putative trehalose synthase"
/protein_id="AAG45375.1"
/db_xref="GI:12018046"
/translation="MEYVTRFSGEGKRLDYREIIGDEALEAKAKAENLQKSPYHV
NSTSFGGVAELIHLNVLPLMRDVGIDARFVLEGNEEFNVKSFINALQKKEILKT
EEMKRLYLEINKNAMEDIDLQFDVYLHDPPAPLIEFEKROPIMVICHILDSPLN
LEFMKFLQFVEKYRIFHMEYVOEDLNQKRVIMPISIPLSKNNELSESELCT
TLERPDVDEPRLITQVARKDEPMKGVEDIDVYRKVEKIPPEOLLIVMAHDEDEG
WLYPEKTLKIGEDDIDKVLNTLGVHAREVNAFOASDVITLQMSIRREGFGLVIEAM
MKKEPYGKAVGSGIKIQIVDGKTFGLVKVNDALIENTLLEHKDYAOMGKNAKERI
KENFTIHLERLIDLNSF"
6772..7788
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45376.1"
/db_xref="GI:12018047"
/translation="MEIPEISHALSELGFTKRYELLTWTLVVGSPSTAKEISTKSGI
PYNRVATISLKLGEVFEIGTPKVVAASPRIAFPEKKELEIMKLEIENNV
KKEORPAIMRSRDEATEMRESLYSAKNVIVYTPPEFEETREDILKLEGGVTV
VSLYIDKIPDLSEFKGKNFYRQPIFKLNLGLMDGKAVYVITQNTPTPSISPPSKS
TYPEIIFSOYSLIETFEKSTLEKETIGNPKDIRFPAEMHADVFNKMLNRYAEI

```

```

TCKNLESGRLFTLGRVVGTYLSLREAVNNHILETENGCVKGMFAVIEDYESTEIK
FMG"
7794..8912
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport ATP-hydrolyzing
protein"
/protein_id="AAG45377.1"
/db_xref="GI:12018048"
/translation="MAGVRLVDMKVFGEVTAVERELSEVKKDGFEMILLGSGCGKTT
TLRMAGLEEPSRGQIYIGDRVADPEKGIIVPPDRDIAMVFOSYALVPHFTVYDNI
APFLKRYPROEITQORREVAELGLPELNRKRELSSGQROVALAGRAVYRPQV
FLMDEPLSLDKLQRMAKELKQROLGVYTIIVHDQVEAMTGDRLAAMNGVL
QOQVSGDEYVDPAWPAAGFISGSPMFLDAIIVDEGVDGEEFLKLLPQGFVGLG
ELGVGREVIFGIRPEDLYDAFARVARGEMLVRAVEIYENLGSERIIVHLRVGVT
FYAFRESRVREGVVDVFDMMKIHIFDKTGAIF"
8959..10461
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45378.1"
/db_xref="GI:12018049"
/translation="MRMCVLAVTTFDPRTALSVREKALKKHNKLEKSSN
LEFVFNELKRYEALREKENGVDKDEEVAKASRLISRDISGIIIGLHMTESNLY
TLARETNKPLLVDADDDPAMAGSTCITSVGSLSMESSNVYHALHVRKGDVERKA
NARAEVAKSLSKSLLMGAYTTGMEHMDLRLKRFIDPLLDQYTIKRRADS
MSDKERRRVEFEFDMENKAKVYDNDKMLPEVRLQAIYVLAQIDIMKYEBOETV
AVSICQPELSIFGVCTACLIPLRPNADMGKRPVPAIDQEGGVKTISSALLFYL
SGKPLPFDIKYVDELVLIANGCSSLYAKLSNPENNELKETTIOGOCQKSGAL
TYRPTPKLTVARLIRKGEYLLFLAGVETIEISKRLKMGOWMPTAIRPLDK
EKLINMGANHLSAVGDTEIEFEVARMWGIKANVRKKEVENFTLSL"
10486..12399
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45379.1"
/db_xref="GI:12018050"
/translation="MSVILSYNGAFVTRONGDMKDGXYFIPTDRFLKGYKLKD
NSVLGSSODGPRKAYSHSIEDEVALLRKRIKNNMAYRELOFYNNSKRVHREVE
YSFKPIEDIEFVRKFGQKIRRIKSSLDGDNESYTGKQIKRNLIKIKTMREKG
LAFAITTEPLEKELVLYEFIRKISKEGELITLTKPLLPVASTNLMDRVERA
IEDLALATLYWGVGWPPAGIPYVACPGRDSITISWLLPYPQYAGCTLKFAKIQ
GKKFPLNDEEBEGKIPHEFREGELSHSKMPAPYAVLADPLVILAAEYLRMTGD
RELIRLKNPLTAAVEMLIRLDECGGYIRYGGLLANOGMDSKEGIPTBEGPTPK
PVALVEYGVAYKALMDAASIGLTLSDPRMLKKEAELKRRNKDFMCDGFALADG
DNVPSKYVSSNNGHLLFTGIAHOKIAERLFEEDMPSGMCIRTISSEKAYNPSYH
NSTIAPHDAVTAIGLASTIGEKAKMLAEALFTAKFDPSNQLPELRSGLSEYPLL
CPRANAPOAMSAASVAFALTALLGLAKRELTVSPLPRNFKVSALVRRKRAYLIES
QEKEVVRPEERDI"
12350..13339
/feature="conserved hypothetical protein"

```

alignment_scores: Quality: Length: 334

Percent Similarity: Ratio: 3.046 Gaps: 7

Percent Identity: 44.611

Align seg 1/1 to: AF307052 from: 1 to: 18001

15 AspValProValIleGluAsnLeuAsnIleThrIleProlysglyserle 31

15043 GATATTCCTCTTCAGAG.....GCCGAACCGGGAGAGGTT 15077

31 upheThrLeuLeuGluIylaseGlyCysGlyLysThrThrLeuLeuArgm 48

15078 TTTTACCTTCTGCGCCGAGTGCTGTGGAAGACAAACACTTGAA 15127

48 etIlealaGlypheAsnserIleGluIylGlyupheTyYrpheAspasp 64

```

15128 TAAATGCTGCTTTGAGAACGACGACGAGAGATATACCTTATGAT 15177
      |||||||||||:.....: ||:.....:|||||||||
      65 ThrLysIleAsnAsnMetGluProSerLysArgAsnIleGlyMetValph 81
      |||:.....: |||
      15178 ACCGTTATGAACGATGTTCTCCATATGAAGAGACATAGGATGCTCTT 15227
      81 eGlnAsnTyrAlaIlePheProHisLeuThrValAlaArgAspAsnValAla 98
      |||||||:.....: |||||||
      15228 TCAGATATATGCTCTCTTCCACATATGACGTTTATGCAACATCTCT 15277
      98 heGlyLeuMetGlnLysLysValProLysGluGluLeuIleGlnGlnThr 114
      |||||||:.....: |||||||
      15278 TCGGTTTAAAGCTCAGAAAGCTCTCAAGAGAGATTTAAAGACGAGTC 15327
      115 AsnLysTyrLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysTr 131
      |||||||:.....: |||
      15328 TCATGGGCTCTTGAGCTTGTGGGTTTGAAGGATTTGAAGACCGATTACC 15377
      131 cAspLysLeuSerGlyGlyGlnGlnGlnArgValThrIleAlaCysAla 148
      ||:.....: |||
      15378 CGAGCAGCTGAGCGGTGGTCACACAGACAGATTCACATGGCAGAGGCTT 15427
      148 euaIaValaAsnProSerValLeuLeuMetAspGluProLeuSerAsnLeu 164
      ||:.....: |||
      15428 TCGTATAGAGCCCGACGCTTGTGCTTTAGATGAACCTCTTCCATCTC 15477
      165 GluAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGlnIleGlnH 181
      ||:.....: |||
      15478 GATGGAAAAATVAGAGAAAGGCTTAGGGGAGAAATTAAGATCCAAAA 15527
      181 sGluValGlyIleThrThrValTyrValThrHisAspGlnGluAla 198
      ||:.....: |||
      15528 AGAGCTTGAATTAACATCAATCTATGATACGCATGTCAGGAGAACTA 15577
      198 eAlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGlnGln 214
      |||||||:.....: |||
      15578 TGGCAATAGTAGTAGATAGACCGGTTATGAGCGTTGAGAAAGTAGACAG 15627
      215 IleGlyArgProLysGluLeuTyrHisLysProAlaAsnGluPheVal 231
      ||:.....: |||
      15628 GTTGGAAACCATGTGATCTCTACTACCAACCAAAAAAAGAGTTGTGAC 15677
      231 aThrPheIleGlyArgThrAsnIleIleProAlaAsnLeuGluLysArg 248
      ||:.....: |||
      15678 CAGGTTTATAGCTCTTCTAATCTTTAGAGGTTAAACGACAGAGGAA 15727
      248 eAspGlyAlaTyrIle.....ValPheSerAspGly 258
      ||:.....: |||
      15728 AA.....GCATATGTTGGGAACCTTTCCTTGAAGTTAAACGAGAGG 15771
      259 TyrAlaLeuArgMetProAlaLeuAspGlnValGluGlnAlaIleH 275
      ||:.....: |||
      15772 .....AGACTAA 15779
      275 sValSerIleArgProGluGluPheIleLysAspGluSerGlyAspIleG 292
      ||:.....: |||
      15780 GATATTTCTTACGCCGAGAGC...GTTATATCGAAGAGGAGAGC... 15822
      292 LngLysThrIleArgAspSerValTyrIleGlyLeuAsnThrAspTyrPhe 308
      ||:.....: |||
      15823 .....ATGGAGAAATAGTGCACATCGAA 15846
      309 IleGluThrGlyPheAlaSerLysIleGlnValSerGluGluSerThrPh 325
      ||:.....: |||
      15847 CTGCTCCCGGCT..... 15858
      325 eGluGluAspLeuGlnLysGlyAsnArgIleArgLeuArgIleAsnThrG 342
      ||:.....: |||
      15859 .....AGAAATACGGCTCAAGATTGATGTGG 15883
      342 In 342
      :
  
```

```

15884 AG 15885
seq_name: gb_ba:AF307053
seq_documentation_block:
LOCUS      AF307053      17854 bp      DNA      linear      BCT 12-FEB-2001
DEFINITION Thermococcus litoralis sugar kinase, trehalose/maltose binding
              protein (male), trehalose/maltose transport inner membrane protein
              (malG), trehalose/maltose transport inner membrane protein (malG),
              putative trehalose synthase, trehalose/maltose transport
              ATP-hydrolyzing protein, putative sulfate transport system permease
              putative, putative sugar-binding transport ATP-binding protein, and
              unknown genes.
ACCESSION  AF307053
VERSION    AF307053.1 GI:12018057
KEYWORDS   1
SOURCE     Thermococcus litoralis.
           Archaea: Euryarchaeota: Thermococci: Thermococcales;
           Thermococcaceae: Thermococcus.
REFERENCE  1 (bases 1 to 17854)
           Diruggiero, D., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J.,
           Horlacher, R., Robb, F.T., Boos, W. and Weiss, R.B.
           Evidence of recent lateral gene transfer among hyperthermophilic
           archaea
           Mol. Microbiol. 38 (4), 684-693 (2000)
JOURNAL    20566786
MEDLINE    11115105
PUBMED     2 (bases 1 to 17854)
REFERENCE  Dunn, D.M.
           Direct Submission
           Submitted (20-SEP-2000) Human Genetics, University of Utah, Rm 308,
           BRB, 20 S. 2030 E., Salt Lake City, UT 84112, USA
FEATURES   location/Qualifiers
           1..17854
           /organism="Thermococcus litoralis"
           /db_xref="taxon:2265"
           844..1161
           /note="conserved hypothetical protein"
           /codon_start=1
           /transl_table=1
           /product="unknown"
           /protein_id="AAG45386.1"
           /db_xref="GI:12018058"
           /translation="MKYAVYLFPLOSFLYLSASAGSERYMLLIGLNPFLAMCI
           VKDERRAIKITIYAKGVDFPALMLMGALFOSLNAIDLALIDLILGLFKKEPT
           ES"
           complement(1145..2083)
           /codon_start=1
           /transl_table=1
           /product="sugar kinase"
           /protein_id="AAG45387.1"
           /db_xref="GI:12018059"
           /translation="MIVAIGEILIDFIAKEEGKLVDREFEKGAPANVVGLRL
           GAKSALISKVDGDFELIEELKEVERKRIIDKTKHGIYFVLGAKKEPILY
           DGVAIVNLKREELQMDPMDAELHRSYVFAFRPSSTYEVYRAVKGVPISYDN
           IRDLWRGKREKMLDIIEALKLADIVKIDGELYLNKNGIALEDNPAVALTRGA
           EGSTIHKDIRVDVSPKVEVDVTGADAFMALNLSTFMGLDILLESKEELKL
           GSFANLVAAALSTTKGAMSVPSLEEVILKHKRFSLP"
           2317..3669
           /gene="male"
           2317..3669
           /gene="male"
           /codon_start=1
           /transl_table=1
           /product="trehalose/maltose binding protein"
           /protein_id="AAG45388.1"
           /db_xref="GI:12018060"
           /translation="MNYKRYVLGLFLVGYGIAYVAGSGCTGGQOTSTVSTPIETSLQ
           GKIVFAVGAPNEIEYWKGVIAEFERKYPGVYELKQATDTEGRRLIDLNALNGKSS
           DPVFLMDVAMLGFIASGMLPELDVVDVOKRDLVSFQSVINLADQGGKLYALPV
           YIDAGLLYRKDLLEKYGYSKRPETMQLVEMAKIQSGRETNPNWGFVMOGKQVE
  
```

gene
3711..4603
/gene="malp"
CDS
3711..4598
/gene="malp"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane protein"
/protein_id="AAG45389.1"
/db_xref="GI:12018061"
/translation="MDNLTSTKLYREAKLGYLMLPLTVLVFIIIPVAGTEWISL
HRYAIEPERKPEVGRNLYRLVLSAREFWSTFVSPFSVLETTIGLSPALINE
RLKGGVRAIVLIPMAVPTIISARTWELMKNYSGLFNMTISITIGSPVNMGLPIS
AFRAIVADVKRTTPPTMLLLAGLQALIPQDLYEALIDGASMEERESTIDPLKPV
LVALITFTIDALRFEDIIIVLTGGPGGATTSISLAFNYSNIGDVGIGSAISILTF
VLVLSFTIVLYLKVGFRFD"

gene
4600..5436
/gene="malg"
CDS
4600..5436
/gene="malg"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane protein"
/protein_id="AAG45390.1"
/db_xref="GI:12018062"
/translation="WREEVLRKRLIIIGAILMAIICLPETIMTVISFAEDPTFLGSP
LVEKSTLENVRLSDPTLHPAYLKNSSIIIASLVLTIVSISLAAYASRIEFG
RLIIPFVLGSMFPOISLVGLFEFIEKLGWNTYQALFPYVAMTPIELMILSY
FSQPKDLDEAMIDGASRIKTLTIIILPSAPALSPALLVTFIAFNEFMALFTT
DHRATVPVGIATLQGVGHEIIPWGSVMAASVISTIPLYIMALLQKIYVSGITAGALK
GE"

CDS
5441..6679
/codon_start=1
/transl_table=11
/product="putative trehalose synthase"
/protein_id="AAG45391.1"
/db_xref="GI:12018063"
/translation="WYEYTRFGSGKRLIEDYREIIGDALNAIKAKANLKDQSVVH
NSTSGGGVAEILHLVPLMRDVGIDAMFVIEGNEFWNTKSFHNLQGNKELRT
EEMKLYLEIKKNNAEDIDLQFDVGLIHDPAPALIEFYEKROPVIRCHIDSDPN
LEFMKLFQVEYKEDRYIFHMEEYVOEDLNQEKVIMPISIDIPSEKMEISESILK
TLERPDVPERPIITOVARPDPMKGVFPIDIVYRKVEKIEPVOLLIGVMAHDDPEG
MIFEKTLRKIGEDDIKVLNLTNGVNAAREVNAORASDYLQMSIRGPGILTYTEAM
WKEKPVGRAVGGIKLQIVDGRTGFLVADVNDALDKTLVLLHDKDVAQEMKNAERI
KENFITRKHLERYGLDLSN"

CDS
6712..7728
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45392.1"
/db_xref="GI:12018064"
/translation="MEIPPEISHALSEIGFTKREYITLWTLIVVGPSTAKESITRSGI
PYNRYDRTISSILKRGVTEIEGPKPYAAVSPRAEFREKLEEDIMKKEILENNY
KKEGORAIMRSRSDAEIEMFRESIYSAKNEVIYVPSSEFETIRELLITLLEGVY
VSLYIDKIPDLSEERKGNFTVROTYKNHLIGMDGEEVATIQATDSIGPVSFKS
TYPEIFESQYSLIIEIFKESYLEKEIIGNPDIIRFPAFVADVFKRNKLKRNVIYAEI
TGMNLSGRLETLTGRVVGTYLSLREAVNNHILTEENGCVAVGVGFVIEDYESTEIK
FIMG"

CDS
7734..8852
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport ATP-hydrolyzing protein"
/protein_id="AAG45393.1"
/db_xref="GI:12018065"
/translation="MAGVRLVDVWKVGEVAVRRLSLFVADGDEPMILGPSGCKTT

TLRMAGLEEPSRQOIYIGRLVADPEKGIFFPPKDRDIAMVFOVALIYPMYVYDNI
AFPIKLRRVPOEIDQREVAELGLTELLNRPRRELISGQQRORVALGRAIVRKPOV
FLMEPIISLDAKLEIDVRMARAEKALQRLQATTYIVTHDQYEAATMGRIAYNRNGL
QOVSQDEYVYKPAINTFVAGFISGPPNMFDAITVEDQFVGEFRELKILPDEQEVIG
ELGVGDEVYIGIRPEQIDYDAMPARVRPGRNLRVAYEIVYENIGSERIVLRRVGAT
FVGAFRSRSRRVREGVEVDVFMKKIHFDKTTGKAIF"

CDS
8899..10401
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45394.1"
/db_xref="GI:12018066"
/translation="MPWCVLAVITFTDPRTALSYEREKALLEKHNLIKELKSSN
LEVPDVALCKEYALERGENGVDDKREYKASRLIGSDISGIIIGLMTESNLV
TLAKETNRPLLLIADDDPAMAGSTCIYSVASLMESSVYNHALHVRALKGDVEKVKY
MARAEAVSKLSKSLSLIMGAPYTLGMEHLMDLPLRKRFIGDFLIDQYIIIRADS
MISPEKEIRVREEPDMLEKAKYNNDKMKLTPLVLRQAIYIYLAADIKMEYQDEY
AVSJKOPELSEIYGVTRCLIPALPPPAADKKGKPPVPATCEBDYKGTISALLFTL
SGKPPLEFDIYVDDELVLNCGASSIYAKLTSEENLEKETTIOGCGGKSGGAL
TYRTPPTKIVARLIRKGEYLLYFLAEGEYLTLEESKELMKGRQMPHVAIRNPLDK
EKLPMGNHLISAVPGDFTIEIFVARMGCIKAVNLRKKEVENFLTSL"

CDS
10426..12339
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45395.1"
/db_xref="GI:12018067"
/translation="MSVILSYNGAFVATRONDKMDKGDGFIYFDTRPLKGVKLKLOK
NSVILGSDQCPKRAKYSHFSIEDVEVLTLRKEIKLKNAGYRELOPYNTSKRYLNRV
YSFVPIEDIEYVRKFGQKIRRIKSSLSGDENYSTGKRIKRNLIKTNMTETEG
LAFAPILTEPLEKLEFLIEIPKISKBELEFLIEKPLLNNVSTNLTWLDVFERA
IEDTALTAAYNYEMVFPAGIPIYACFGFRSITSMFLPYPYQVAGEGLKFPFAKIQ
GKNNLEEEEPGKIPIHEERFGESSHSSKMPFAYGVGDPLPLYIAEYLRMTGD
RELIERLKPNLTAAVEMTLRLDGGGYIRKXGILLNCGKDSKEGIPREEGPTP
PVALYVQGYAKALMDAASGLTSLDPKMLKRAEELKRFNPDPCDGYAALDLC
DNVPSKAVSSNMGHLFTGIAEHQEKIARLEFEDMFSGKIRTLSSKERAYNPFSTH
NGSIMPNDNAVIAIGLASIGERKARMAEAIFFYAFVLPSPQSLPELPSGLESYPLL
CPRANAPQDMSASVFFALITLALGIAKEKELTVSPLLPKNKVSALVFRKKAIVIES
QEKRVPRDEERDI"

CDS
12290..13279
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45396.1"
/db_xref="GI:12018068"
/translation="MNLKRRRLDLKSRIPEKIKIKGNKIDIKPPKIPPIPSR
EGFDSRNTYNPAVITKEDRIYMLYRAESKEDKLTGRGLAISTDGINFPRIPEETIEP
EYKWEVSVEDPRLIVKGYTYMTYTGDTARLCLATSNMLTMRKHGIFEGFEY
KNNIKGWTLSGALIPKLEKGNFKGNLMEFGSNNIMAVSKDLVNMVEYKKEVLSR
REGEDVILVPEPPLPEFTSYGIALIYSAGRYDKLKYKGVVIFPKREYDKVAART
ETPLMVPEYEMELYGHVNNVYFVGWHEHKKILLYGGADRHIGLAWYTDL"

CDS
13303..14934
/note="abc transporter"
/codon_start=1
/transl_table=11
/product="putative sulfate transport system permease

alignment_scores:
Quality: 683.50 Length: 334
Ratio: 3.024 Gaps: 7
Percent similarity: 67.665 Percent identity: 44.611

alignment_block:
US-09-769-787-162 x AF307053 ..
Align seg 1/1 to: AF307053 from: 1 to: 17854
15 AspValProValIleGluAsnLeuAsnIleThrIleProLysGlySerIle 31
||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
14982 GATATTCCCTCCCTAGAG.....GCAAAACCGGAGACAGTT 15016

```

31 uPheThrLeuGluGlyAlaSerGlyCysGlyThrThrLeuValArgm 48
   ::::::::::::::::::::|::::::::::::::::::|
15017 TTTAACTGCTCGCGCCGAGTGGCTGTGGAAGACACACACTTAGAA 15066
   ::::::::::::::::::::|::::::::::::::::::|
48 eCileAlaGlyPheAsnSerIleGluGlyGlyLeuPheTyrPheAsp 64
   ::::::::::::::::::::|::::::::::::::::::|
15067 TAAAGCTGGTTTGAAGCCAGCAAGGAGACATATCTTATGAT 15116
   ::::::::::::::::::::|::::::::::::::::::|
65 ThrLysIleAsnMetGluProSerLysArgAsnIleGlyMetAlaPh 81
   ::::::::::::::::::::|::::::::::::::::::|
15117 ACCGTTATGAACGATGTTCTCCATATGAAGAACATAGGATGCTCTT 15166
   ::::::::::::::::::::|::::::::::::::::::|
81 eGlnAsnTyrAlaIlePheProHisLeuThrValArgAspAsnValAla 98
   ::::::::::::::::::::|::::::::::::::::::|
15167 TCAGGATTTAGCTCTCTTCCGATATGACGTTTATGCAACATCTCCT 15216
   ::::::::::::::::::::|::::::::::::::::::|
98 heGlyLeuMetGlnLysLysValProLysGluGluLeuIleGlnGlnThr 114
   ::::::::::::::::::::|::::::::::::::::::|
15217 TCGGTTTAAAGCTCAGAAAGCTCTCTAAGAGACATTAAGACGAGTC 15266
   ::::::::::::::::::::|::::::::::::::::::|
115 AsnLysTyrLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysPr 131
   ::::::::::::::::::::|::::::::::::::::::|
15267 TCATGGCTCTTGAGCTTGTGGTTGAAGGATTTGAAGACCGTTACCC 15316
   ::::::::::::::::::::|::::::::::::::::::|
131 oAspLysLeuSerGlyGlyGlnGlnArgValThrLeuAlaCysAlaL 148
   ::::::::::::::::::::|::::::::::::::::::|
15317 CGAGCAGCTGACGCGCGCTCAGCAGAGAGAGTGTGACGGAAGGCTT 15366
   ::::::::::::::::::::|::::::::::::::::::|
148 eAlaValAsnProSerValLeuLeuMetAspGluProLeuSerAsnLeu 164
   ::::::::::::::::::::|::::::::::::::::::|
15367 TGTGATATACAAACCCAGCTTTGCTTTAGATGAACCCCTTCCATCTC 15416
   ::::::::::::::::::::|::::::::::::::::::|
165 GluAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnH 181
   ::::::::::::::::::::|::::::::::::::::::|
15417 GATGCAAAATATAGAGAAAGGCTTAGGGAGAAATTAAGATCCAAAA 15466
   ::::::::::::::::::::|::::::::::::::::::|
181 sGluValGlyIleThrThrValTyrValThrHisAspGlnGluAlaAl 198
   ::::::::::::::::::::|::::::::::::::::::|
15467 AAGAGCTGGAATACACTACATCTACATGCAATGTCAGAGAGAACTA 15516
   ::::::::::::::::::::|::::::::::::::::::|
198 eAlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGlnGln 214
   ::::::::::::::::::::|::::::::::::::::::|
15517 TGGCATATAGTATAGATAGCCGTTATGAGCGTTGGAAGATAGAGCAG 15566
   ::::::::::::::::::::|::::::::::::::::::|
215 IleGlyArgProLysGluLeuTyrHisLysProAlaAsnGluPheValAl 231
   ::::::::::::::::::::|::::::::::::::::::|
15567 ATTGGGAATCCATGTGATCTCTACTACCAACCCAAAAAATGAGTTGTAGC 15616
   ::::::::::::::::::::|::::::::::::::::::|
231 aThrPheIleGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgS 248
   ::::::::::::::::::::|::::::::::::::::::|
15617 CAGGTTTATAGCTCTTCTAACAATTTTAGAGCTTAACGACAGAGAGGAA 15666
   ::::::::::::::::::::|::::::::::::::::::|
248 eAspGlyValAlaTyrIle.....ValPheSerAspGly 258
   ::::::::::::::::::::|::::::::::::::::::|
15667 AA.....GCATATGTTGGGAACCTTGTCTTGAAGTTAAACGGAAGG 15710
   ::::::::::::::::::::|::::::::::::::::::|
259 TyrAlaLeuArgMetProAlaLeuAspGlnValGluGlnAlaIleH 275
   ::::::::::::::::::::|::::::::::::::::::|
15711 .....AGAGTAA 15718
   ::::::::::::::::::::|::::::::::::::::::|
275 sValSerIleArgProGluGluPheIleLysAspGluSerGlyAspIleG 292
   ::::::::::::::::::::|::::::::::::::::::|
15719 GATATTCTTTAGGCCGAGAGAC...GTTATATCGAAGAGGAGAC.... 15761
   ::::::::::::::::::::|::::::::::::::::::|
292 LuGlyThrIleArgAspSerValTyrLeuGlyLeuAsnThrAspTyrPhe 308
   ::::::::::::::::::::|::::::::::::::::::|
15762 .....ATGGGAGAATAAGTAGCTACGAA 15785
   ::::::::::::::::::::|::::::::::::::::::|
309 IleGluThrGlyPheAlaSerLysIleGlnValSerGluGluSerThrPh 325
   ::::::::::::::::::::|::::::::::::::::::|
15786 GTCCTCCCGGT..... 15797

```

```

325 eGluGluAspLeuGlnLysGlyAsnArgIleArgLeuArgIleAsnThrG 342
   ::::::::::::::::::::|::::::::::::::::::|
15798 .....AGATACGGCTCAAGATTGATGTGG 15822
   ::::::::::::::::::::|::::::::::::::::::|
342 In 342
   ::::::::::::::::::::|::::::::::::::::::|
15823 AG 15824
   ::::::::::::::::::::|::::::::::::::::::|
seq_name: gb_ba:SME591790
seq_documentation_block:
LOCUS SME591790 323450 bp DNA linear BCT 16-AUG-2001
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.
ACCESSION AL591790 AL591688
VERSION AL591790.1 GI:15075230
KEYWORDS
SOURCE
ORGANISM
Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
1 (bases 1 to 323450)
REFERENCE
AUTHORS Capela,D., Barloy-Hubler,F., Guzy,J., Bothé,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetalle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
From the Cover: Analysis of the chromosome sequence of the legume
symbiont Sinorhizobium meliloti strain 1021
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9877-9882 (2001)
11481430
PUBMED 2 (bases 1 to 323450)
Guzy,J.
REFERENCE
AUTHORS Direct Submission
Submitted (26-JUL-2001) Guzy J., Submitted on behalf of the MELILO
EU Consortium
COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Génétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unité de Biochimie physiologique,
Université catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté
des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Guzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
location/Qualifiers
1. 323450
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
155. 1519
/gene="SMC01575"
155. 1519
/gene="SMC01575"
/function="miscellaneous: hypothetical/partial homology"
/feature="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=1
/evidence="not experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC6787.1"
/db_xref="GI:15075231"
/translation="MTPATRTGSLFPAALAMELAPANQDAYGYWGCGDVMYLT
PEPRDRIYIGEAHVHAMRDGRGTIVDPNGNTVATVPNDGVRQRRREGRDGY
PEPRDRIYGYSEPEFTGAIPEYRDIAPAVEREDLPNSLPSLSREAAVDPQYDPP

```

LAQPMPTVYTKSRALIAQLVFLDREGSPGVIDGKMSNTKAIEMOQATGET
LDPNNTEDILERLFRNGGLPITTYITADADAQPVASIPEDYARKADLPHSFTSVT
EMLGEKFMHDEARLYLRELPVDFSIPTTITKVVNGPKKGVARIVADKSKQVLA
DEAGKLIAVPESTIGSDTSPSGTVHERIAFDGTYNPKINFOOGANDIILLOP
GPNPGVTGWIATLSKPTYGIGHTPEPSIKGKQSHGCVRLTNWDATELGKMWSTGTV
EPVD"

gene
1707..2996
/gene="SMC01576"
1707..2996
/gene="SMC01576"
/function="small molecule metabolism"
/note="product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC46788.1"
/db_xref="GI:15075232"
/translation="MATERIALPMLHATAPAPRPAPLAGDLTVETAVGGGTLSA
ALHAEFTGRTAVTEARMIKGGSGRNVGLVAGWQPPDILATLGAAGRIIDEL
GDGSEFYDLVAKKHCIECEYANGTLHNSVAGELKEIREPQAKKAGAPVEUISAE
KAHALSGAEGFTGALLDRRAGTIPLAYARGLARALAGAEIFDTPLAASRGDL
WNLKPGCTVAVARHIALATNAGSLVTGVPKMEYKQELTILPYEQFATNPDPNAR
ILPEROGAMDTGLVMTSEFMDRONRLIFGSIQRLAIAAGTHRAFAARSVRKLFPYIG
DREFHMDGRIGMTNNLPAHWLAPVNSISGNGTGVAGTVEGRALARHVTGDT
SAIPLAETPVTPDPRTLSAFYHAGAKHIDKR"

gene
complement(3024..3740)
/gene="SMC01579"
complement(3024..3740)
/gene="SMC01579"
/function="miscellaneous: hypothetical/global homology"
/note="product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46789.1"
/db_xref="GI:15075233"
/translation="MAVITFANAKGAGCTTAALITSTELAROGNRVVVALDPORMI
TMSSEVGRVANLEVISHTVPASLPCHIRELKGADFIYDLAAGAKKDALVATGLSD
HYLIPVOCAMDARGAVOILELIRIGEKARIRINHSVYLVTVNSLVTTALQTTKAL
LASRGVSALDPTVERAVARETFECGGTLLQMDMPRVNSLNDKARENAYALAEVONL
PYTARRALMSRLRSALPRAA"

gene
complement(3949..5094)
/gene="SMC01582"
complement(3949..5094)
/gene="SMC01582"
/EC_number="1.1.1.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE ALCOHOL DEHYDROGENASE PROTEIN"
/protein_id="CAC46790.1"
/db_xref="GI:15075234"
/translation="MTIRANVSYPAAKFGAGRIKELADHCKALGIRKPLVYDRGLA
PMAIQALDILEAGLIGRAIFADVPDNPBNDRIEAGYKAPFADGGHGVAFGGSGSL
DLGCVAFMAGOTRPVMDPEIDIGMWTASVEGIAPIYAVPTTACTGSEVGRASITIN
SASHYKVIHFHKPLGVITCDPELTGMPVYITTAGTMDFAHGLEAVISFPHKMS
AGIALEGRLVKEYLPRAVYKDGADLEARANNMSAAGVAVFOKGLAIHLSHPVGA

IYNTHGNTAVNAPVPLRNRNSAIEKIGRAAYLGIAGFDDGYDYVLRLRELVG
PDKLSALGVGDRIDEMAKEMAIVPTAGNPEVETLDAERKLFACET"

gene
complement(5091..6476)
/gene="SMC01588"
complement(5091..6476)
/gene="SMC01588"
/EC_number="1.2.1.-"
/function="small molecule metabolism"
/note="product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE ALDEHYDE DEHYDROGENASE PROTEIN"
/protein_id="CAC46791.1"
/db_xref="GI:15075235"
/translation="MTIRCVSPVDEGVYERPAIPLERMAQVAHAHARLAKANARRP
LDERVKVLAVARLERNVDEVPDLAQMKRPVYGGEEFGFERNRSYVASIADAL
KPLVVEESDREFERRIAREPHGVNVEIAPWNPYMTAINTVAPALMAGNTVILKASOT
ILVGERVRAFTIENGVPADVFONFLDHTTAAIIAASKFDFINFTGSEGRSIERA
AAGFTGTGLGEGKPGVYVMDADIDAAVDTLMDGATVNSGCCGGERIYVHESLY
DAFERKSAVYSNKRLGNPLDPEPTLGGMAKKRRATYRNOYADAISGARAILDPKL
FPQDDGAYVLAPQVLYVDVDSHMEFERRETFEPVAGIMVKSDAEILEMDSRGTLTA
SLWTKDARARIRIGDLETGTVMNRADYLDPLALCMVGKVGKGTGRGSLSYDGFHNLTR
PKSYLLKKVTA"

gene
complement(6674..8038)
/gene="SMC01594"
complement(6674..8038)
/gene="SMC01594"
/function="miscellaneous: hypothetical/global homology"
/note="product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46792.1"
/db_xref="GI:15075236"
/translation="MSYFEEELKEDVAAGRIDTVLACVDMQGRIMGRFRAEYVES
AMKTEHSCNYLADDMEMETVPGYKATSEMGYGDYTKPDLSTLRIRPMEGTALYL
CDMLDHDHVAEHPSPRAILKQVARELAMEGFAVMASELEFFLEDSYDARLSGR
DLOLASGVNEGYHIFOTKEDVMAIRNGIOGAGIPENSKGASAGOEINRYAD
ALTMAADRATITKNGCKETANORGRAITFLAMWNSAAGSSSHIHOSIKSGENPLRF
DKNGOYGESELMRHHVAGQLAHASVYVFLAPYINSTRFPAAGTAFPKATWSKDNR
AGYRLCGESKAIRIECRVGSGLNPIYLAFAALIAAGIAGIENKMELEAPVGDAYOG
KEVERIPIHTLREAGALSGLSKMLRAAFEEVEVDHYVHAWEEOOEYDRRYVDWEVARG
FERA"

gene
complement(8131..9678)
/gene="SMC01597"
complement(8131..9678)
/gene="SMC01597"
/function="cell processes: transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1

alignment_scores:
Quality: 678.50 Length: 377
Ratio: 2.590 Gaps: 5
Percent Similarity: 69.496 Percent Identity: 39.523

alignment_block:
US-09-769-787-162 x SME591790

Align seg 1/1 to: SME591790 from: 1 to: 323450

```

1 MetSerGluLeuLeuIleLeuAlaLysLysIleThrHisValPr 17
15447 ATGGCATTCCTGCATATTCACAAATTCCTCGGCCCGCTCA 14696
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
14697 GGTGCGCATATTCGACATGGGGATCCAGACAGGCGAGTGTCTCTCT 14746
34 eubLeuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
14747 TTCTCGGCCCTCCGCGTGGCGACAGCATCTCGGCATGATTGCC 14796
51 GlyPheAsnSerIleGluGlyGlyPheThrPheAspAspThrLysII 67
14797 GGTTCGAGACCCGTCGCGGCTCATCTTCATCGACGCAAGCA 14846
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGluAsnT 84
14847 GGTGCGCTGAGCCGACACGACGCAATATCGCATGTGTTCAGGCGCT 14896
84 yAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
14897 ATGCGGCTTTCCCAATGATGTCACAGCAATGTCGCGCTCGGCGCTC 14946
101 MetGluLysValProLysGluGluLeuIleGlnGlnThrAsnLysTy 117
14947 AAGTCGCGCGCGCTCCAGACGAGATCGAGCAGCGGTAAAGCAT 14996
117 rIeuGluLeuMetGlnIleAlaGlnTyAlaAspArgLysProAspLysL 134
14997 GCTCGACATGATCAAGCTAAGCATCTGGCGGACCGCTTCCCTATCAAT 15046
134 eusErGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
15047 TGTCCGCGGACAGACAGCGGTGGCTTCCGCGCGCTCGGCGCTC 15096
151 AspProSerValLeuLeuMetAspGluProLeuSerAsnLeuAlaTy 167
15097 AACCCGAGGTGCTGCTCCGACGAGCGCGCTCTCCGCGCTCATGCAA 15146
167 sIeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
15147 GATCCGATCTCCCTCGGAGAGATCCGCCAGATTCACACAGCATCG 15196
184 yIleThrThrValTyValThrHisAspGlnGluAlaMetAlaIle 200
15197 GCATCACACAGGTATTCGTAGCAGCAGACAGAGAGGCGCATGTGCATC 15246
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
15247 TCCGACCGGATCGTCGATGAATTCGCGCGCGCATGATGATCGGCAC 15296
217 gProLysGluLeuThrHisLysProAlaAsnGluPheValAlaThrPheT 234
15297 CCTTTCGAGATCTACACACCCCGGAGACGCTTGTCCTCTCTCG 15346
234 lGluIleGlyThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
15347 TCGGACGCTCAACATCATCGAGGCAAGGTGGCCGATCCCGAAGCGGC 15396
251 AlATyTilLeValPheSerAspGlyTyAlaLeuArgMetProAlaLeuAs 267
15397 GAGTGCACCATCGCGGCGGCGCATCTCACTAAGAGAGCGATTGCCGG 15446
267 gAlnValGluGluAlaIleHisValSerIleArgProGlu.....G 282
15447 GGCMAAAGGCGGACAGCATCTCGCTGCGCTCAGGCCGGAAGCAGCGCT 15496
282 luhelIeLysAspGluSerGlyAsp.....IleGluGlyThrIleArg 296
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

15497 CGATCGCCGAGGGCGCCAGGGCGCATACCGCATCTCCGGGCGAAGTGTGA 15546
297 AspSerValTyIreGluY..... 302
15547 TCGAGGAGCTTTCGCTCCGTCATCCGACCAACCTCGGGTGGCAA 15596
303 .....LeuAsnThrAspTyPheIleGluThrGlyPheAlaSerLysI 317
15597 AGACATCATCTCTTCGATATGTTCACAGATCCGGGCATGACCCACCGG 15646
317 lGluValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 333
15647 TC.....GCCGCGAG 15657
334 ArgIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAla.. 349
15658 AGGCTACCGCTGCGTTTCGCGCAAGACCTGCTCATCGCGAATA 15707
350 .AspGlySerGluAsnLeuIleLysGlyVal 359
15708 GACGCGCCGGAATGAATTTGTCAGGGCGGTG 15738

seq_name: gb_ba:AE008131
seq_documentation_block:
LOCUS AE008131 11311 bp DNA linear BCT 18-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 189
of 254 of the complete sequence.
ACCESSION AE008131 AE007869
VERSION AE008131.1 GI:15157288
KEYWORDS
SOURCE
ORGANISM
Agrobacterium tumefaciens str. C58 (Cereon).
Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 11311)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 11311)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
source
1..11311
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/strain="C58"
/db_xref="taxon:181661"
77..1072
/gene="AGR_C_3892"
77..1072
/gene="AGR_C_3892"
/note="probable permease of ABC transporter PA3252
(imported) - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=1
/product="AGR_C_3892p"
/protein_id="AAK87895.1"
/db_xref="GI:15157289"
/translation="MKRTKIRADSVYLTGRCARPLETDRENGYWGCGMPENNTG
GPSDAGRKLPLNMIGVPPFAVFMFLILPMPKLVIGFQGTDSFTLDNAGLFTSS
ILAAVWISIKISLASAALGCLIGFAVAANAVALGLPQRIKRPDLTFSGVASQFAGVPL
AFARFATIGPVLGTLVFLKQIGIDRLIGNIIISFPGILVTYLFPIPLMILITPFA
LDGLKREMRBAEILIGATGLOLYRMVAAPILPILGMSLLFANAGCAVATAATG
SSINIVPLLEPAQIRGDVLTGSRPHLGVALAFGMIVLTGIANNLYIMLARSRWLK"
1060..1851
/gene="AGR_C_3893"
1060..1851
/gene="AGR_C_3893"
/note="probable ABC transport system permease Cj0731

```



```
{imported} - Campylobacter jejuni (strain NCTC 11668)"
/codon_start=1
/transl_table=11
/product="AGR_C_3893p"
/protein_id="AK87896.1"
/db_xref="GI:15157280"
/translacion="MAEMKRFEMWALITFGLLYFALPLIGMTNFSIKMRGEYSPDAY
GKVLDPREFQETFSYVMALFTIVFGVLLVPTAYVRLKPLGRPIEFTITLPLV
IPAVIVFGYIRLNTSSMLPIGTFTGLMLMFGYATLALPYMYRAVDGIRITDV
STLEAASIGAGMTITLSRIILPNVLNVLSCAFLLTFAIVIGSEFTMALINPARGP
YMOLGANRAYEPAALAVISFGITWGCIGLIQIVSRVOKGAPPKA"
1865..2923
/gene="AGR_C_3896"
/gene="AGR_C_3896"
1865..2923
/gene="AGR_C_3896"
/feature="ABC transporter, ATP-binding protein VCA0602
(imported) - Vibrio cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_C_3896p"
/protein_id="AK87897.1"
/db_xref="GI:15157291"
/translacion="MSFLTLNKKISFGSVQVYVDFNMAIEKGEFVFLGSGCGKTT
VLIMAGFEIPGSGTVINGKQDTLPAPNOMTGMFOAYALPNNMYENVAFGKY
AGKPAEIDARVEMLIQILHEHLADRYPTOMSGGQOYALALAPPEVLLDEP
LSALDAKIRSLREETROIQKLGITTTITVTHQEBALISDRIVMNGSRADQIGSP
FDLYNKPATRVASFEVGLNLIDATVVDSSSNTIRVEQOITLPKRVNANGKRIILA
LRPEAGSLGADGVAISGLVYSQFASVIRTRLDGSLTSPDMFDPGTAPPAV
GERVTLKFAAGDMLIKD"
2988..4061
/gene="AGR_C_3897"
/gene="AGR_C_3897"
2988..4061
/gene="AGR_C_3897"
/feature="ALCOHOL DEHYDROGENASE (ADH-HT)"
/codon_start=1
/transl_table=11
/product="AGR_C_3897p"
/protein_id="AK87898.1"
/db_xref="GI:15157292"
/translacion="NNLRINDEAMRALLYERGETPVVASLPDPAPSGGVYAVKA
TGLCRSDMHWGMHDIDIRLPHYGHEPAGVISAGRNTRYETGTGRVVPYPSGCGH
CHECRSGNOVCETOFQPGFTHMGSAEYVAIDYAOQNLVHLPESYATMAELGCRF
ATSERAVTDGRLKGGEMLAHGGVGSLAINIGALAOVVAIDYADKELTALV
GAVATTINRSADVAAVNDITGGGAHVSDALMGHPQTCNSISNLRGRHVOGLM
LAADHAMPALPMARVIAHLEIETGSHGQAMRYEDMLAMIESGRLAEPKILGRHSILE
AAVALPGMRFOFSGSIIIDRE"
complement(4089..4730)
/gene="AGR_C_3900"
/complement(4089..4730)
/gene="AGR_C_3900"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3900p"
/protein_id="AK87899.1"
/db_xref="GI:15157293"
/translacion="MSYTAIFSAGALFSYIFASVAERKDRLEFAISESGGETIALVY
DLPTNPAPLAEGNTLDLGLFNSDSGSGYLVYGDRTYRLSDADVGRKSRIGEDPTIO
HRLERARERKONERNMLAFETIVLVGFLVYIKGDAYRMEKRLITMLVDTDKGAA
EHRDPLKARKKRLVRYQSRPRTDVAASPAPEPPVASEPTGTATAFGRRT"
5063..5839
/gene="AGR_C_3902"
/gene="AGR_C_3902"
5063..5839
/gene="AGR_C_3902"
/feature="AP001508) BH0315
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/product="AGR_C_3902p"
/protein_id="AK87900.1"
/db_xref="GI:15157294"
/translacion="MLFLDSVAMKEKPEGRREYPRSPVPMRHLERLEFRTPTTFPAGN
NSGKSTLLEGLAAGMTAYAIIGNHGDVKGDLVLIQHAETVAKAEYFARKRYPKIRMLR
```

```
AEADVGIYRQNEMLNIDFERWEREALKKGDEPPEETPEITFNGIYRNNSLDRSHEG
FLDIMHRLRHAGLYFLUDEPESPLPOKOLELAALIRQAAADSGGOLLATHTSPVLLAI
PEATYVFEDEDDITERLYDELEENIAFLRRFLDRPSKYLD"
5918..7249
/gene="AGR_C_3904"
/gene="AGR_C_3904"
5918..7249
/gene="AGR_C_3904"
/feature="probable MFS transporter PA5370 (imported) -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_3904p"
/protein_id="AK87901.1"
/db_xref="GI:15157295"
/translacion="MSPTTTPAARPLNSODPRTGLSALGALFEYDFITFEPATV
IGHLEFPPEMPMLVMIQFEGIFAGYIVLPAGGIVLAYGGRKRYKFAFASILLMA
LSLIGMALPMTYATGVNAPILLITLRMOAGAIAGEVGAMTFVAREPVRGVLAC
GRTSGISGIMUGSLIARINSIEPPEVAVARIPILGIVGLIAYVIRRLKEE
PTIYEMKRSKSLTDKPLGLVYKHHRCVITISALTYVLSAIVYTTIMTPTIQKL
YGVTPQALAGTSFGTLIFGVIIAGALIDVSGIFSGMSIFFGIATFTFYSYAG
TSLETFEVLGVGMAGVAVPVVWAPASVFSGLSEFAYVNSAVAFGLGTPIG
VTTALAINPMAMWLVFIATFLAFAIGLYLHLRGSSESVHGLEELALRS"
complement(7323..11102)
/gene="AGR_C_3907"
/complement(7323..11102)
/gene="AGR_C_3907"
/feature="methionine synthase PA1843 (imported) - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_3907p"
/protein_id="AK87902.1"
/db_xref="GI:15157296"
/translacion="MPVFDLDFPEGAKRDGAEIFKALRDASERLLIDGAMGTQIQ
GIGPDEDHRRGDRFTGCACHOKNDRLILTPQPDALIEEIRHAYMAAGAILETNFSS
THIAQADYEMENAVYDLNREGAAYVRRRAQADRIILLETITDILAKAAIPACERTEAK
VNNPGTRAVSFDDLRIATGEQIDGLIDGADRIILLETITDILAKAAIPACERTEAK
GIRLPMYISGTTITDLSGRTLSGQTPSAFENSVRHNPFITIGLALGADAPMLHQL
SDVADTFVCAVYPNAGLPNEFGYDETPKMAHVOVGFDGLVNVIVGGCGSTPBIIR
ATAEAVKDVKPREIPEHKRPMSLSGLEPVLTKDIPFVWGERTRVWTSARPKLITA
GAVAVTLIPMEEEKRLNGSSERSAAGVLMATVKGVDVHDKGNIVYACNAYETID
LSVMPPTTKIILETALAIEKVDYIGSLITPSIDENVYHAAEBEROGFDRIILGATIT
SVHTAVKTHPRYEQOAIYVTDASRANGVSVALISEQRPATIDGIRIETAKVAEH
ARNERKORPLPSRARENHAKIDWSYSVYKQFGTGFETVILDEISRYIDWTPPF
QWELKGRPALIIDEKOGKEARQIYADAOAMQARIIEKKMPRAVAVIGFWPNAVGD
DIRLFTDECKREELATFTFLROOLSKRDRPNVALSDFYAPVDSVADVAGGFVYTAG
IREVIAIAERFERANDYSSILYKALADRPAEAFERMEHRYRKEKMGVAPDEALADD
LIGEYAGIRPARPGYPAQDPDTEKKTIFLALDNTAACTELTESYAMPGSSVSLYI
GHPESTYFVANVARDQVLDYARKKDMPTVEVRNIGVPLNTVPTNGEKIDSA"
BASE COUNT 2245 a 3522 g 3155 g 2389 t
ORIGIN
alignment_scores:
Quality: 671.00 Length: 331
Ratio: 2.750 Gaps: 4
Percent Similarity: 73.716 Percent Identity: 44.109
alignment block:
US-09-765-787-162 x AE008131 ..
Align seg 1/1 to: AE008131 from: 1 to: 11311
1 MetsergiuilelysileleasnalalyslylletyrhtiaspValPr 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1865 ATGAGCTTTTGGACACTTACGACATCAAGAAATCCTTCGGCTCGTGA 1914
```

```

17  oValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1915  GAGTCGTCACATTTTCACATGGCATCGAAGGCGAATGCTGCTCCT 1964
   * 34  euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1965  TTCTCGGCGCATCGGGCTGTGGCAAGCCACGCTGCTGCCCATGATCGCC 2014
   51  GlyPheAsnSerIleGlyGlyGlyLysThrPheAspAspThrLysIL 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2015  GGTTCGAAATCCCGACCGCGGATCATGCTCATCAACGCGCAAGACCA 2064
   67  eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGluAsnT 84
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2065  GAGACGCGCTCAGACCGACGCGCATATCGGCATGCTTCCAGGCGCT 2114
   84  yAlaIlePheProHISLeuThrValArgAspAsnValAlaPheGlyLeu 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2115  ATGCGCGTTCGCGACATGATGCTATGAAAAACGTCGCTCGGTCTG 2164
   101  MetGluLysValProLysGluGluIleGlnIleThrAsnLysTY 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2165  AAAGTCGCGCGCAACCAACCAAGCCGAGATCGAGCCCGCGTGAAGAAAT 2214
   117  rLeuGluLeuMetGlnIleAlaGlnTYrAlaAspArgLysProAspLysL 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2215  GCTCCAGCTCATCATCTGACATCGGCGGAGCGCTACCCCTATACAGA 2264
   134  euserGlyGlyGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaVal 150
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2265  TGTCTGGCGGTTCAGACACGCGGTGCGCTCCCGCGCGCTTCGCGCG 2314
   151  AsnProSerValLeuMetAspGluProLeuSerAsnLeuGluAlaVal 167
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2315  AAGCGGAATGCTCTGCTGGAGAACCGCTCTCCCGCTCGAGCGCAA 2364
   167  sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHISGluValG 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2365  GATCCGCGGTCTCGCTGAGGAATCCCGCATCCAGATCCAGACAACTCG 2414
   184  lyIleThrValTYrValThrHISAspGlnGluGluAlaMetAlaIle 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2415  GCATCACACCATTTCTGACCCACATCAGGAAGGCGCTGTCATCA 2464
   201  SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyYar 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2465  TCCGACCGCATCTCGTGAATGAGCGCGCTGCCGACAGATCGGCTTC 2514
   217  gProLysGluLeuTYrHISLysProAlaAsnGluPheValAlaThrPheI 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2515  GCCCTCGATATCTACAAACACCGCGCTGCTGCGCTCCTTCG 2564
   234  leGlyArgThrAsnIleIleProAlaAsnLeuGlyLysArgSerAspGly 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2565  TCGGTACGTTGAACCTCATCGACGCCACGCTGTGATTCCTCATTCAT 2614
   251  AlATyrIleValPheSerAspGlyTYrAlaLeuArgMetProAlaLeuAs 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2615  ACCATCCGTGTCGAGAGACAGCATCACCTCGCGCAAGCG..GTGGA 2661
   267  pGlnValGluGluGlnAlaIleHISValSerIleArgProGlu.....G 282
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2662  TGCAGCGCAATGGCGAATAACGCTGCGCATTCGCGCGGAGACAGATT 2711
   282  lPheIleLysAspGluSerGlyAsp.....IleGlyGlyThrIleArg 296
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2712  CCCCTCGACAGATGCGAAGGCGATGTTCATTCGCGCTCGTCACCC 2761
   297  AspSerValTYrLeuGly..... 302
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2762  TCCAGCGAGTTCCTCGGCTCGCTCATCCGACCGCGCTTATGTCGGCG 2811
   303  .....LeuAsnThrAspTYrPheIleGluThrGlyPheAla 314

```

```

||||: ||| ||| ::: ||| |||
2812  TTCGACGCTGCTCCTGACATGTTCAACGACCCCGCACGCGC 2854
seq_name: gp_ba:AE009165
seq_documentation_block:
LOCUS      AE009165                12416 bp    DNA        linear    BCT 20-DEC-2001
DEFINITION      Agrobacterium tumefaciens strain C58 circular chromosome, section
                  191 of 256 of the complete sequence.
ACCESSION      AE009165 AE008688
VERSION        AE009165.1 GI:17740611
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmeri,A., Raymond,C., Rouse,G.,
Saenphummachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.,
and Nester,E.W.
The Genome of the Natural Genetic Engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
11743193
2 (bases 1 to 12416)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmeri,A., Raymond,C., Rouse,G.,
Saenphummachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.,
and Nester,E.W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA
FEATURES
source
location/Qualifiers
1..12416
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
73..1179
/gene="Atu2147"
73..1179
/gene="Atu2147"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate binding protein"
/protein_id="AAL4137.1"
/db_xref="GI:17740612"
/translation="MISCRRLATTTALVASTAIAAPSASIELIAARKESMLTIV
ALPHDMCGYGVIAIAFKAKIPEITVNEINPDAGSADVEAYKANKDKGPAPVDVY
GIAEPQKASGLLOPYKVSWTAEIPDVKADGKWDGYGVSMFVNKOLVKNVPK
WDADLIPKEYGQVLAADPPASNOAILIIVAGLSTGAKGKSGEAGEGLVFPADLTK
AGNPLVIGKAGTIAOGATPIITIVADYVALSMKTKLNDNPPTPEVAVPEKGLVACYVQ
GISAAYAPHNAAKIMMEHLVSDGGLGKXGCHPIRNNAMVAKGKVPQGLIDSLPRA
AAIEKAIPTLIEVDANKAAVATGGCKDYGVGNV"
1383..2267
/gene="Atu2148"
1383..2267
CDS

```

gene
CDS
/gene="Atu2148"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein"
/protein_id="AAU43138.1"
/db_xref="GI:17740613"
/translation="MPSNNTGSPDAGRKPLNMIGVPEFAVMFLPLPTMKIYIG
AFQDTGSEFTLNDIAGFTSSIIAAWISIKISIASAGCLGFAVAAYVGGIDP
RIKGPLTFSGVSOFAGVPLAFATFATGVLGTLPLFKTQIDRLGPNLISFM
GLVYTLFQIPLMITLITTPALDGLKREKREAEITGATGQIWRVAVAPFILPDLG
TMSLTFANAFGAVATAIALTGSSLNITVPLLEFQIINGDVAGNPHLGVALAFMIVLTG
IANALYIMLARERERWLK"
/gene="Atu2149"
/gene="Atu2149"
/gene="Atu2149"
/gene="Atu2149"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein"
/protein_id="AAU43139.1"
/db_xref="GI:17740614"
/translation="MKRFFAWGALTFGLYFALPLIGMTNFSIKMRGEYSEDAVKV
LTPDROETFSYSVMALEFTYVGVLLVPTAYVWRLKLGRLPYIEFTLPLVPA
IYIVFGYIRLYNYSWMLPLTGFTNLIMLEGALPYMYRAVDGTGRTDVSFL
TEAOSLIGAGWTLSRIILPNYLVAVISGAPTFPAVIGEFMALNLRPAFGPIYQ
LIGANRAYEPALAVISFGITWGCLGILQVSRYOKGAPPKA"
/gene="Atu2150"
/gene="Atu2150"
/gene="Atu2150"
/gene="Atu2150"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein"
/protein_id="AAU43140.1"
/db_xref="GI:17740615"
/translation="MSFTLSNITKKSFGSVQVVDHFNMAIEGEEVSYLSPGCCGKT
VLRMIAGFEIPTGGSIVNGKQDTTLRPQNRNGVQVAFLEPNNAVYENAVFGKLV
AGKPKAEIDARVKEMLQILHEHLADRYPOMSGQOQVALALAPYEPVILLDEP
LSADAKIRVSLKEEIRLOKQIGITTYIETHQEEALISDRIVVMNGRADQIDSP
FDIYNKPARPVASFYGTNLNLDATVVDSSNTRIGBEOITLPRKPVDAANGKRTILA
LREASGLSDAKGVYISGLVTSQFLGSVITRDLGSGTSLPFMFNDPGIAPAPV
GERVTLKFAAGDLMLIKD"
/gene="Atu2151"
/gene="Atu2151"
/gene="Atu2151"
/gene="Atu2151"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="alcohol dehydrogenase"
/protein_id="AAU43141.1"
/db_xref="GI:17740616"
/translation="MRALYEREGETPVVASLPDPAISDGCVIAVAKATGLCRSDMHG
WMGHDIIDILPHYPGHEFAGVISAGRNVTREFTGDRVIVPEVSGGHEGCSGQO
VCETFOPGFTHMGSPAEYVAIDYQONLVHLPESVSATAGLCGRFATSPFAVTD
GRKLGKEMLVHGGCGVLSAIGAGLAQVVAIDIAEDKLEIAROLGATATINRS
VADVAAVADITDGGCAHYSVDALGHPTQCNCSNRRGRHYQVGLMADNAPAIIP
MARVIAHELEITGSHGQAMRTEDMLAMTESGRLAEKLIGRHISLTAFAVALPGMDR
FQESGISIIDREF"
/gene="Atu2152"
/gene="Atu2152"
/gene="Atu2152"
/gene="Atu2152"
/note="putative; ORF located using Glimmer"

gene
CDS
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAU43142.1"
/db_xref="GI:17740617"
/translation="MSYTPALFSSAGALFSYIFASVAEAKRDRLEPAISEGCEETIAY
DLPNTPAFIABEANTLDIGLNSDSGSGVLYIGDRYTRLSDDVGRSLRGLFDDTQ
HLEERAAREKQNERNNLAIEFTLVAVGFPLVVKGLDAVRWRLTLMVDTDKGOAA
EFDHDLAKRKRRLVROYSRPRTDSVAASPVPYEAASPTGTARAGRRRK"
/gene="Atu2153"
/gene="Atu2153"
/gene="Atu2153"
/gene="Atu2153"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein"
/protein_id="AAU43143.1"
/db_xref="GI:17740618"
/translation="MLFDSVAMKEKPEGREGPEPSPVAMRHLERLEFPTPTFPAGN
NSCKSTLLEGIAACMTATAIGNHGDYKGLYLOHAETVAKAFYARRKYPKIRMEFLR
ADVLGYIRQENALDDERWERKALKGEDPPEETPTFRTGTYRNNSLDRSGEG
FLDIMHRLHAGLYFLDEPESPLSPQKLELAIIROAADSGGLIATHSVLLAI
PATIYVPEDECDITERLYDELINIAFLRFRFLDRPSKYLDD"
/gene="Atu2154"
/gene="Atu2154"
/gene="Atu2154"
/gene="Atu2154"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="MFS permease[proline/deltae]"
/protein_id="AAU43144.1"
/db_xref="GI:17740619"
/translation="MSPTTPAARPLNSODEFTGLSALGALFEYDITLFFEPFATV
IGHLEFPPEPMDLVWIOFGIIFAGYLVPRAGGLVLAHYGRYARRKRYAFSILMA
LSTIGMALMPRYATIGVAPILLITLIRMOGALAGTEGVAFTFAEHVPERVGLAC
GFLTSGLSRGIMUGSLIAFATNSLTPEVYAGYARRIPRLLAGVGLIAYIRPLEE
PTIYEMKRSKSLTDKLPGLVLRKHRRKQVITISALIVLSAIVITLMTATFLQKL
XYVTPPQALAGTSFGLTLFEGVILAGALIDVVGSGIIFMGASIEFGIATFTFYSYAG
TSLEMFVYGVWGSLVSGAGAVPYVMVAPFASVFSGLSFAVNVSAVFEGLPIG
VTTAALINMAWLVYFAPAFPAIIGLYLHRLGSEVSHVIEBLAARS"
/gene="Atu2155"
/gene="Atu2155"
/gene="Atu2155"
/gene="Atu2155"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="methionine synthase"
/protein_id="AAU43145.1"
/db_xref="GI:17740620"
/translation="MDDLFCEPGEKARRGAEIFKALRDASERILLIDGAMGTQIOLG
GDEDEHFRGDRITGCACGKGNNDLILITQPAIDIEIHRYAMACADILERTFTSSSTR

alignment_scores:
Quality: 671.00 Length: 331
Ratio: 2.750 Gaps: 4
Percent Similarity: 73.716 Percent Identity: 44.109

alignment_block:
US-09-769-787-162 x AE009165 ..
Align seg 1/1 to: AE009165 from: 1 to: 12416

1 MetcercutlletysllleasAlalyalslleyrhtsasValPr 17
||||| : : : ||| ||||| : : : |||||

OM of: US-09-769-787-162 to: N.Geneseq_032802.* out_format : pfs

Date: Sep 12, 2002 12:18 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed_p2n.model -DEV=xlh
-O=qm2.1/USPTO.spool/US09769787/rnat.12092002.084403.29379/app_query.fasta.1.429
-DB=N.Geneseq_032802 -QPMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=6.000 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsnum2
-TRANS=human40.cdl -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09769787_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-769-787-162
Query length: 363
Database: N.Geneseq_032802.*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 209.620000
```

Score list:

```
Sequence      Strd Orig      ZScore      EScore len 1 Documentation
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA05429 + 1843.00 3512.65 2.6e-187 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV2260 - 1843.00 3489.67 5.0e-186 95
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AA30946 + 1003.00 1906.61 7.5e-98 69
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA528538 + 669.00 1225.84 6.2e-60 3
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA53495 + 651.50 1227.22 5.2e-60 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA86431 - 651.00 1165.72 1.4e-56 3
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA30248 + 650.50 1154.64 5.7e-56 91
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA42063 - 650.50 1147.26 1.9e-55 18
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA86431 + 648.00 1159.97 2.9e-56 3
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA52319 + 638.50 1203.12 2.4e-54 3
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA41224 + 630.00 1125.45 2.1e-54 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA54628 + 628.00 1155.18 5.4e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA94377 + 627.50 1182.81 1.6e-57 9
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546237 + 627.00 1155.60 5.1e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546235 + 627.00 1155.60 5.1e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546274 + 627.00 1155.60 5.1e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546278 + 627.00 1155.60 5.1e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA313035 - 626.00 1149.86 1.1e-55 16
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA52850 + 623.50 1174.01 4.8e-57 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA54442 - 623.50 1164.09 1.7e-56 2
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA48815 + 620.50 1168.11 1.0e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA48816 + 620.50 1168.11 1.0e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA52318 - 620.00 1129.56 5.8e-54 3
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA303041 + 619.50 1082.78 5.8e-52 29
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA313068 + 619.00 1144.47 2.1e-55 79
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA32020 - 619.00 1127.65 1.8e-54 38
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV2318 - 617.50 1146.76 1.6e-55 48
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAV52318 - 616.00 1130.71 1.2e-54 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV52355 + 612.50 1141.45 3.1e-55 32
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA81457 - 612.50 1112.56 1.3e-53 49
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA81457 - 612.50 1091.89 1.8e-52 34
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA81460 + 612.50 1091.89 1.8e-52 34
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AA81490 + 610.00 1076.94 1.2e-51 14
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV44359 + 610.00 1144.66 2.1e-55 15
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA211330 + 605.00 1136.57 5.8e-55 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AA296314 + 605.00 1137.94 4.9e-55 11
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA055583 + 605.00 1127.58 1.9e-54 30
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV52271 - 601.50 1125.74 2.3e-54 19
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AA296313 + 601.00 1111.71 1.4e-53 66
```

```
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA232025 + 599.00 1103.82 3.9e-53
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA030082 + 599.00 1103.82 3.9e-53
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA03041 - 596.50 1038.67 1.7e-49
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA66498 + 594.50 1117.21 7.0e-54
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA68528 + 594.50 1057.37 1.5e-50
```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA05429

seq_documentation_block:

ID AAA05429 standard; DNA; 1092 BP.

AC AAA05429;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae nucleotide sequence ID211 - 4127.2.

KM Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;

KW pneumococcal disease; ds.

OS Streptococcus pneumoniae.

PN WO200006737-A2.

PD 10-FEB-2000.

PE 27-JUL-1999; 99WO-GB02451.

PR 27-JUL-1998; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Gilbert CPG, Hansbro PW;

XX WPI; 2000-195300/17.

XX P-PSDB; AAY81674.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of

XX PT pneumococcal diseases and for screening agents capable of antagonizing

XX PT or inhibiting expression of the protein

XX Claim 6; Page 92; 108pp; English.

XX AAY1501 to AAY81679 represent specifically claimed protein sequences

XX isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent

XX specifically claimed nucleotide sequences isolated from S. pneumoniae.

XX The sequences have antibacterial and antiinflammatory properties.

XX CC The protein sequences, and fragments of them, are useful as immunogens

XX and/or antigens. The nucleotide sequences can be used in vaccines and in

XX diagnostic assays. The proteins and nucleotides can be useful for the

XX detection and diagnosis of S. pneumoniae. The protein sequences are also

XX useful for screening an agent capable of antagonizing, inhibiting or

XX interfering with the function or expression of the proteins in which the

XX agent is useful for treatment or prophylaxis of S. pneumoniae infection

XX CC and meningitis. AAA05591 to AAA05614 represent primers used in the

XX exemplification of the present invention.

SQ Sequence 1092 BP; 361 A; 226 C; 219 G; 286 T; 0 other;

alignment_scores:

Quality: 1843.00 Length: 363

Ratio: 5.077 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-769-787-162 x AAA05429 ..

Align seg 1/1 to: AAA05429 from: 1 to: 1092

1 MetSerGluIleuLysIleIleAsnAlaIaLysIleTyhrHisAspValPr 17

```

|||||
1 ATGAGTAGATCAAAATTATTATACGCCAAAAAATCTACCGATGTCCTCC 50
17 oval11Eg1uasnleuasn1lethr1lePro1ySg1Ser1eupher1r1 34
51 TGTATTGAGAAATTTGAAACATTACAAATTCAAAAGAAAGTCTTTTACC 100
34 euleug1yalaSer1yCysg1yLysThr1leu1euar1eal1eal1e 50
101 TTTCTGGAGCTTCAGAGATGTGGAAAACGACCTTCTTCGTATGATGCA 150
51 GlypheasnSer1leg1uGlyGlyuPheryrPheaspasph1r1y11 67
151 GGTTCACACAGTATCGAAGGTGAGAAATTTTCACTTCATGATACAAAAT 200
67 easnAsm1e1uProser1yAr1asn1leg1Me1yAl1Pheg1Inas1n 84
201 CAATTAATATGAAACCCAGCAACCAATATCGGATGTTTCCAAAAC 250
84 yrala1lePhePro1leuThr1yAla1r1gAs1nVala1Ala1Pheg1yLeu 100
251 ACGTATTTTCCACATTTGACTGCCAGACACGTTGCTTTGGTCTT 300
101 Met1In1yS1yVal1Pro1ySg1u1eul1leg1In1thr1Asn1yS1y 117
301 ATGCAAAAGAGGTTCCAAAAGAAATTTGATTCACACGACCAACAAATA 350
117 r1eug1u1eum1e1n1le1ag1n1yAla1s1p1r1g1yS1yPro1yS1y 134
351 TCTTGAACCTCATGCAAAATGCTCAATATCGGATCGAAAGCCCATTAAC 400
134 eus1er1yGly1yGln1yGln1yAla1r1yAla1r1eua1a1Cys1a1leua1a1 150
401 TCACTGTGTGACACAAACACGCTGACCTGGCATGGCTTACGCGTT 450
151 Asn1Proser1Val1leu1eum1e1As1p1u1Pro1euser1asn1eug1u1a1y 167
451 AATCCAAGTGTCTCTCATGAGACGACCACTTGTATCTGGAGGCCAA 500
167 sleu1r1eua1s1p1e1ar1g1Ina1le1ar1g1u1leg1In1s1g1u1a1g 184
501 ACTTCGCTGTGATATGCGTCAAGCATCCGAAATCCAAACACCAAGTGG 550
184 l1y1lethr1r1Val1y1r1Val1Thr1s1as1p1g1u1g1u1a1Me1a1le 200
551 GAATTAACAAGCTTTATGTAACCCACACCAAGAAAGCATGGCTATT 600
201 Ser1sp1g1n1le1a1Val1Me1y1s1p1yAl1leg1In1leg1yAr 217
601 TCAAGACCAAAATGCTGTATGAAAGATGGGTGATCCAAACAAATCGGCG 650
217 gPro1ySg1u1eul1y1r1s1yS1yPro1a1Asn1g1u1Phe1Vala1lar1Phe1 234
651 ACCAAAAGAACTCTATCATTAACCAAGCTAATGAGTTTGTGCAACCTTTA 700
234 leg1yAr1gThr1Asn1le1lePro1a1asn1eug1u1y1r1gSer1Asp1y 250
701 TCGAGCCCAAAATATATCCCTGCAATCTTGA AAAAGAGGAGCGCG 750
251 Alar1y1r1le1Val1PheSer1Asp1y1r1Ala1eua1r1gMe1Pro1a1eua1 267
751 GCTTATATCTCTTTTTCAGATGCTATGCTTCGAAATGCGACGCTTTGA 800
267 gAl1Vala1g1u1g1Ina1le1s1Val1Ser1le1ar1gPro1g1u1g1u1e1 284
801 TCAAGTTGAGAGACACACTATTCTATGTAAGCATTCGTCGCCAAGATTTA 850
284 le1y1As1p1u1Ser1y1Asp1le1g1u1g1y1r1le1ar1g1s1er1Val1y1r 300
851 TCAAAAGATGATCTGAGATATTTAAGAACTATTTAGAGATAGCGCTTAT 900
301 leu1g1y1eua1sn1thr1Asp1y1r1Phe1leg1u1Thr1Gly1Phe1a1s1er1yS1 317
|||||

```

```

901 CTTGGACTAAATACGATTAATTCATTGAGACAGGTTTCCCTCAAAAT 950
317 eg1Val1Ser1g1u1Ser1Thr1Phe1g1u1g1As1p1eug1In1ySg1yAsn1 334
951 TCAAGTTAGTGAAGATCACTTTTGAAGAAGATCTCAAAAAGGCATC 1000
334 rgl1eAr1g1eua1r1le1asn1Thr1Gln1y1eua1sn1lePheSer1Ala1s1p 350
1001 GTATTCCTCTACGAATCAATACGCAAAAATTTAAACATCTTTCTGCAGAT 1050
351 GlySer1Gln1asn1eul1e1yS1yVala1s1n1s1g1yThr 363
1051 GGTTCACAAAACCTGATAAAAGAGTCAACCATGGAACG 1089
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV52260
seq_documentation_block:
ID AAV52260 standard; DNA; 9578 BP.
XX
AC AAV52260;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae; genome fragment SEQ ID NO:127.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN W09818931-A2.
XX
PD 07-MAY-1998.
XX
PE 30-OCT-1997; 97MO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences; - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 876-882; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridize to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and

```


CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
 xx Sequence 9578 BP; 2886 A; 1653 C; 2116 G; 2923 T; 0 other;
 SQ

alignment_scores:

Quality: 1843.00 Length: 363
 Ratio: 5.077 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-769-787-162 x AAV52260/rev ..

Align seg 1/1 to reverse of: AAV52260 from: 1 to: 9578

```

1 MetSerGluIleLysIleIleAsnAlaLysIleTyrHisAspValP 17
|||||
8112 ATGAGTGAATTTGACATTTACATTCACAAAGAGAGTCTTAAACC 8063
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerIleuPheThrL 34
|||||
8062 TGTATTGGAATTTGACATTTACATTCACAAAGAGAGTCTTAAACC 8013
34 euLeuGlyAlaSerGlyCysGlyLysThrThrLeuAsnArgMetIleAla 50
|||||
8012 TTCTTGAGCTTCAGATGTGGGAAACGACCTTCTGATGATGCA 7963
51 GlyPheAsnSerIleGluGlyGluPheTyrPheAspAspThrLysI 67
|||||
7962 GGTTCACAGATTCAGAGTGGAGATTTTACTGATGATCAAAAAAT 7913
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGluAsn 84
|||||
7912 CAATATATATGAGACCAACGCAATATCGGATGCTTTTCCAAACT 7863
84 yAlaIlePheProHisLeuThrValArgAspAsnAlaPheGlyLeu 100
|||||
7862 ACGCTATTTTCCACATTTGACTGTCGAGACACGTTGCTTGGCTT 7813
101 MetGluLysLysValProLysGluGluLeuIleGlnGlnThrAsnLys 117
|||||
7812 ATCCAAAGAGGTTCCAAAGAGATTTGATTCACAGCCAAACAGTA 7763
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLys 134
|||||
7762 TCTTGACATATGCAATGCTCAATATGCGGATCGAAGCCCGATTAAC 7713
134 euSerGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
|||||
7712 TCAGTGTGGACAAACAACGTCGACCTTGGCATGCGCTTAAGCGGT 7663
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuAlaLys 167
|||||
7662 AATCCAAAGTGTCTCTCATGGAGCCACTAGTAATCTGGAGGCCAA 7613
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
|||||
7612 ACTTCGTTGGATATGCGTACAGCCATCCGAGAAATCCAAACAGAGTGG 7563
184 lylIleThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200
|||||
7562 GAATTACAACTGTTATGTAAACCCAGCAGCAAGAGCATTGGCATAT 7513
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGly 217
|||||
7512 TCAGACCAAAATTCGTTATGAAGAATGGGATATCCAAATAATCGGCCG 7463
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||
7462 ACCAAAGAACTATATCATTAACAGCTAATGAGTTTGTCGAACCTTTA 7413
234 lGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
|||||

```

```

7412 TCGAGCACAATAATATTCCTGCCAATCTTGAAAAAGGAGCGACGGC 7363
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
|||||
7362 GCTTATATCGCTTTTCAGATGGCTATAGCCCTTCGATCGACACTTTGA 7313
267 pGluValGluGluGlnAlaIleHisValSerIleArgProGluGluPhe 284
|||||
7312 TCAGTGTGAGGACAGCACTATTCATGTAGCATTCCTCCGAAGATTTA 7263
284 lLysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300
|||||
7262 TCAAAGATGAATCTCGAGATATGAAGGAACATATAGATAGCCGTAT 7213
301 LeuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLys 317
|||||
7212 CTGGACTAAATACGAGATTTATTCATTTGACACGAGTTTGCCCAAAAT 7163
317 eGluValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 334
|||||
7162 TCAGTGTAGTGAAGATCAACTTTTGAAGAAGATCTACAAAAAGCAATC 7113
334 rGluLeuArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
|||||
7112 GTATTCGCTACGATCAATACGCAAAATAATTAACATCTTTCTGCAGAT 7063
351 GlySerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363
|||||
7062 GGTTCCAAAACCTGATTAAGAGATCAACCAACGAGAACG 7024
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AAx30946
seq_documentation_block:
ID AAX30946 standard; DNA; 693 BP.
XX
AC AAX30946;
XX
DE 20-MAY-1999 (first entry)
XX
XX Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:223.
DE Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW Streptococcal infection; pneumococcal; ss.
XX
OS Streptococcus pneumoniae.
XX
PN W09737026-A1.
XX
PD 09-OCT-1997.
XX
PF 01-APR-1997; 97WO-US05306.
XX
PR 22-AUG-1996; 96US-0025788.
XX
PP 02-APR-1996; 96US-0014690.
XX
PA (SMTK ) SMITHKLINE BEECHAM CORP.
XX (SMTK ) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
DR WPI: 1997-503111/46.
XX
DR P-PSDB: AAY11367.
XX
PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT vaccines, drug screening, etc
XX
PS Claim 5; Page 185; 354pp; English.
XX
CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,
CC isolated from Streptococcus pneumoniae, can be used in vaccines against

```

CC streptococcal infections and in assays for identifying compounds that
 CC inhibit or activate the activity of the proteins. The antagonists can
 CC be used to treat an individual having need to inhibit a bacterial
 CC protein. Vectors expressing the proteins can be used to induce a
 CC protective immune response in mammals.

XX Sequence 693 BP; 218 A; 155 C; 145 G; 171 T; 4 other;

alignment_scores:

Quality: 1003.00 Length: 231
 Ratio: 4.580 Gaps: 5
 Percent Similarity: 94.805 Percent Identity: 90.909

alignment_block:

US-09-769-787-162 x AAX30946 ..

Align seg 1/1 to: AAX30946 from: 1 to: 693

```

38 SerGlyCysGlyLysThrThrLeuArgMetIleAlaGlyPheAsnSe 54
|||||
3 TCAGGATGTGGGAAAACGACCCCTTCTCGATGATGTCAGGTTTCAACAG 52
|||||
54 rllieguglyglyluphetyrPheaspapThrLysIleAsnAsnMetG 71
|||||
53 TATCAAGATGGAGAAATTTACTTCGATGATACAAAATCAATATATATGG 102
|||||
71 lupProSerLysArgAsnIleGlyMetValPheGlnAsnTYrAlaIlePhe 87
|||||
103 AACCCAGCAACGCAATATCGGCTGTTTCCAAACCTACGCTATTATTC 152
|||||
88 ProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLysL 104
|||||
153 CCACATTGTGACTGTCCGACGACAGCTTTGGTCTTTATGCAAAAGAA 202
|||||
104 sValProLysGlyGluLeuIleGlnGlnThrAsnLysTYrLeuGluLeu 121
|||||
203 GGTTCCAAAGAAAGAAATTTGATTCACACAGACCAACAACTATCTGGACTCA 252
|||||
121 etGlnIleAlaGlnTYrAlaAspArgLysProAspLysLeuSerGlyGly 137
|||||
253 TGCNAATTCGTCATATCGGATCGAAGCCCGATTAACCTCAGTGGTGA 302
|||||
138 GlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSerVa 154
|||||
303 CAACAACAAAGCTGTACCTTGGCATGGCTGTAGCGGTATATCAATGT 352
|||||
154 IleuLeu.MetaspGlu.ProLeuSerAsnLeuGluAlaLysLeuArgLe 170
|||||
353 TCTCTCTCCAGGAGGAGCACTTAGTAATCTGGAGGCCAAACTTCGCTT 402
|||||
170 uaspMetArgGlnAla...IleArgGluIleGln...IleGluValGlyI 185
|||||
403 GGATATGCGTTCAAGCCCATCCGAGAAATCCAAACGAAAGTTGGGGAA 452
|||||
185 lenthThrValTYrValTYrHisAspGlnGluAlaMetAlaIleSer 201
|||||
453 TTAACAAGCTTTATGTAAACCCAGACCAAGAGAGCCATGGCATTTTCA 502
|||||
202 AspGlnIleAla.ValMetLysAspGlyValIleGlnGlnIleGlyArg 218
|||||
503 GACCAAAATTCCTGTATGTGAAGATGGGTGATCCAAACAATCGGCCGAC 552
|||||
218 rOlysgGluLeuTYrHisLysProAlaAsnGluPheValAlaThrPheIle 234
|||||
553 CAAGAAGACTGTATCATAAACACCTATATGATGATGAGCAACCTTATTC 602
|||||
235 GlyArgThrAsnIleLeuProAlaAsnLeuGluLysArgSerAspGlyAl 251
|||||
603 GGAAGCGCAATATTTATCCTGCAATCTTGAAAAAGGAGCGAGCGCGCN 652
|||||
251 aTYrIleValIlePheSerAspGlyTYrAlaLeuArgMet 263
|||||

```

653 TTAATTCGCTCTNTTCAGATGGANANAGCCCTTCGAATG 689

seq_name: /SID1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28538

seq_documentation_block:

ID AAF28538 standard; DNA: 31147 BP.

XX AAF28538;

DT 04-APR-2001 (first entry)

DE Genomic fragment #25.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

OS Moraxella catarrhalis.

PN W0200078968-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000MO-US16649.

PR 18-JUN-1999; 99US-0140121.

PA (INCY-) INCYTE GENOMICS INC.

PI Lagace RE, Patterson C, Berg KL;

DR WPI: 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -

PS Claim 1: Page 207-215; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

SQ Sequence 31147 BP; 8961 A; 6045 C; 6778 G; 9363 T; 0 other;

alignment_scores:

Quality: 669.00 Length: 376
 Ratio: 2.544 Gaps: 7
 Percent Similarity: 69.947 Percent Identity: 40.160

alignment_block:

US-09-769-787-162 x AAF28538 ..

Align seg 1/1 to: AAF28538 from: 1 to: 31147

```

1 MetSerGluIleLysIleIleAsnAlaLysIleTYrHisAspValPr 17
|||||
18353 ATGAGCTATATTTCAATAAACAATGCCACATTAAGTCTTCGTTTCATTGAC 18402
|||||
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
|||||
18403 CGTTATATGATGATTTGAATTAATGTAGAAAGGAGTATGTTACTT 18452
|||||
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
|||||

```

```
18453 TACGTGTCCTTCAGCGTCGGCAAGTCACTTTGTTACGTTGTAATGCC 18502
51 GlyPheAsnSerIleGluGlyGlyIuphetyrPheAspSerThrIys11 67
18503 GGTCTGGAAACATTGATCAGCGGTTCAATTAATTAACAAATCAAGATAT 18552
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnT 84
18553 TACGTATTAAACCCAGAAACGCCGTATGCGCATGATTTCAAAATTT 18602
84 yrfAlaIePheProHisLeuThrValArAspAsnValAlaPheGlyLeu 100
18603 ATGCTCTCTCCCAACATGACAGTGGCAGATATGTAGAGTTGGATTA 18652
101 MetGlnLysLysValProLysGluGluIleGlnGlnThrAsnLysTy 117
18653 AAGATTAAAAAGTCTCTCTTGAGAGAGACTGACTAAAGTCAAAAGCT 18702
117 rLeuGluLeuMetGlnIleAlaGlnIlyrAlaAspArgLysProAspLys 134
18703 GCTTGATTGGTTGAGTTAACTTCTTTGCTCAGCAAAAACGAGATCTT 18752
134 euSerGlyGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
18753 TATCTGGTGAGAAAAACAGCGTTGCACTGGCTAGACATTAGTTATG 18802
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGlnLys 167
18803 GAGCCGATTTGCTTTGTTAGTGGCCATTCCTGCTAGATGCCAA 18852
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184
18853 ACTAGCTAAAGTTTACGATGCAATTTAAGCGCATTCAAAAAGATTGG 18902
184 yLleThrThrValIyValThrHisAspGlnGluGluAlaMetAlaIle 200
18903 GATTAATCACTTTTGTACTCATGATGATGATGATGATGATGATGATG 18952
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGly 217
18953 TCAGACGAGGTGCTTCTGTTAATTAAGGCAAGATTGACACATTCCTC 19002
217 gProLysGluLeuYrHisLysProAlaAsnGluPheValAlaThrPhe 234
19003 GCCAGACACATATATACCTACGCTAACATAGATTACTGCGGGTTTA 19052
234 lGcIyArGThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
19053 TAGGACATTACAAATATA..... 19069
251 AlaTyrlIleValPheSerAspGlyTy.....AlaLeuArgMetProAl 265
19070 .....GGATATTTTGAATCAGTAAATCCAAATC 19098
265 eLeuAspGlnValGluGluGlnAlaIleHisValSerIleArgProGlu 282
19099 TGGCAAGCAACTTTCGATGATGCA.....ATAAGCCCTGA 19136
282 lupheIleLysAsp..GluSerGlyAspIleGluGlyThrIleArgAsp 297
19137 CATATATTGTAGATACGATGATGAGATATTCACAGGCTTATCTTGA 19186
298 SerValTyLysGluLysLeuAsnThrAspTyrrPheIleGluThrGlyPheAl 314
19187 AGAAGCTTACTGGGGAGTGGCTTACCAAGTCAGAACAAATATG 19236
314 aSerLysIleGlnValSerGluLysSerThrPheGlnGluLysAspLeuGln 331
19237 TGATATATTGATGTT.....GATGTAAGTCAACC 19265
331 ysgGlyAsnArgIleArgLeuArgIleAsn..... 340
19266 ATGGCAAAATTTAGCCAAATTAAGTCAATTCAGAGTTTATTAATATC 19315
```

```
341 .....ThrGlnLysLeuAsnIle.....PheSerAlaAspGlySe 352
19316 AAAAGATGAACCAAGTGTATCAACCTAGATTAATTTTCTACCTTAGGATA 19365
352 rGlnAsnLeuIleLysGlyValAsnHis 361
19366 CAAAACACTAGCTACTAAATATGCATCAT 19393
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA53495
seq_documentation_block:
ID AA53495 standard; DNA; 1146 BP.
XX
XX AC AA53495;
XX
XX 13-FEB-2002 (first entry)
XX
XX Haemophilus influenzae DNA for cellular proliferation protein #277.
XX
XX DE Antisense; ds; prokaryotic cellular proliferation gene;
XX
XX KW antibiotic; antibacterial; drug design.
XX
XX OS Haemophilus influenzae.
XX
XX PN W0200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US09180.
XX
XX PR 21-MAR-2000; 2000US-191078P.
XX
XX PR 23-MAY-2000; 2000US-206848P.
XX
XX PR 26-MAY-2000; 2000US-207727P.
XX
XX PR 23-OCT-2000; 2000US-242578P.
XX
XX PR 27-NOV-2000; 2000US-253625P.
XX
XX PR 22-DEC-2000; 2000US-257931P.
XX
XX PR 16-FEB-2001; 2001US-269308P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX
XX PI Yamamoto RT, Xu HH;
XX
XX DR WPI: 2001-611495/70.
XX
XX P-PSDB: AA035636.
XX
XX PT New polynucleotides for the identification and development of
XX
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX PS Claim 27; Seq ID No 7132; 51pp; English.
XX
XX
XX The invention relates to antisense inhibitors of genes essential to
CC
CC prokaryotic cellular proliferation, their use in identifying the
CC
CC genes, their use in the discovery of novel antibiotics, the essential
CC
CC genes themselves and the encoded proteins. The prokaryotes used are
CC
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC
CC invention is also useful for the identification of potential new targets
CC
CC for antibiotic development. The antisense nucleic acids can also be used
CC
CC to identify proteins used in proliferation, to express these proteins,
CC
CC and to obtain antibodies capable of binding to the expressed proteins.
CC
CC The proteins can be used to screen compounds in rational drug discovery
CC
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC
CC for homologous nucleic acids which are required for cell proliferation in
CC
CC a wide variety of organisms. The present sequence encodes an
CC
CC essential prokaryotic cellular proliferation protein.
CC
CC Note: The sequence data for this patent did not form part
CC
CC of the printed specification, but was obtained in electronic
CC
CC format directly from WIPO at
CC
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 1146 BP; 378 A; 203 C; 237 G; 328 T; 0 other;
```


PA (UYJO) UNIV JOHNS HOPKINS.
VY

PR 21-APR-1999; 99FR-0005034.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX WPI, 2001-126236/14.
DR
XX
XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode
PT proteins useful in industry -
PS
XX Claim 1; Page 183-279; 1657Pp; French.
XX
XX The present invention relates to the genomic sequence of *Pyrococcus*
CC *abyssi* and *P. abyssi* proteins (see AAB96053-AAB96842). *P. abyssi* is a
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC vents. The present sequence is a fragment of the genomic sequence of *P.*
CC *abyssi*. The 3' end of this sequence overlaps with the 5' end of AAB41223.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAB75903-AAB75920 and AAB66436.
XX
XX
SQ Sequence 349980 BP; 95167 A; 78444 C; 78870 G; 97499 T; 0 other;

alignment_scores:
Quality: 648.00 Length: 406
Ratio: 2.382 Gaps: 14
Percent Similarity: 66.995 Percent Identity: 39.409

alignment_block:
US-09-769-787-162 x AAF86431 ..

Align seg 1/1 to: AAF86431 from: 1 to: 349980

1 MetSerGluLeuIleuIleAsnAlaLysIleThrHisAspValPr 17
184563 ATGGCCGAACTCAAGCTGATTAACGTGTGAAAAAGTTCGCTGATTTCAC 184612
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
184613 CCGAGTTAAGAGCTAAGCCTTGAGTAAAGATGCGGAGTTTAACTGT 184662
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
184663 TATTAGGGCCAGCTGGGTGTGTGAAGACTACACCCCTTAGATGATGCGC 184712
51 GlyPheAsnSerIleGluGlyGluPheThrPheAspAspThrLysIL 67
184713 GGATGTGAAAGCCAAACATCGCGCAGATATCATGATGATCAAGATTGGT 184762
67 easnAsnMetGlu.....ProSerLysArgAsnIleG 78
184763 TCGTGATCCGTAAGAGGCTATCTTCCTCCCAAGAGAGGAGACTTG 184812
78 lYmeValAlaPheGlnAsnTyraIleAlaIlePheProHisLeuThrValArgAsp 94
184813 CAATGCTTTTCCAAAGTTATGCTTGTGACCCCTCATATGACCGCTACGAC 184862
95 AsnValAlaPheGlyLeuMetGlnLysLysValProLysGluLueuIL 111
184863 AACATAGCCCTTCCCACTAAAGCTAAGAGAGTTCCAAAGCAAGAGANTGA 184912
111 eGlnGlnThrAsnLysTyraLysLeuMetGlnIleAlaGlnTyraLaa 128
184913 CAGAGAGGTAAAGCAAGTTCGAGATGCTTGTGTTGACGACGACTTTAA 184962
128 spArgLysProAspLysLeuSerGlyGlnGlnGlnArgValThrLeu 144

184963 AGAAAAACCTAGAGACTTTCAGGTGGCGAGAGAGAGAGTGCCTT 185012
145 AlAcysAlaLeuAlaValAsnProSerValLeuLeuMetLysPglProle 161
185013 GGAAGGGCCATTATTAGAGACCGAAGGCTTCTCTGATGGATGAGCCTT 185062
161 uSerAsnLeuGluAlaLysLeuArgLeuAspMetArgGlnAlaIleArg 178
185063 AAGTAATCTGGATGCAAAAGTTGAGAGTTAAGATGAGAGCTGACTTAAGA 185112
178 LuIleGlnHisGluValGlyIleThrThrValIleThrHisAspGln 194
185113 AGTTACAGAGACAGTTGGGTGTAAACGATATACCTTACCCAGATCAA 185162
195 GluGluAlaMetAlaIleSerAspGlnIleAlaValMetLysAspGly 211
185163 GTTGAGGCGCATGACAGATGAGATGAGATTGCGTGTGACACAGGGAGA 185212
211 lIleGlnGlnIleGlyArgProLysGluLeuTyraHisLysProAlaAsn 228
185213 ACTCCAGCAAGTTGGAACCCCTGATGATGTATACCAACAGCCCTCAATA 185262
228 lUpheValAlaThrPheIleGlyArg.....ThrAsnIleIleProAla 242
185263 CATTCTGCGCGGGTTTCATGAGGAGCCACCAGATGAACTTCTGACCGCT 185312
243 AsnLeuGluLysArgSerAspGlyAlaTyraIleValPheSerAspGly 259
185313 ACAGATCA.....TCAGACGGG.....TTCTTGATATTG 185344
259 rAlaLeuArgMetProAlaLeu...AspGlnValGlu..... 270
185345 AGAATTAGGTGAACCTCTTAAGATCAATTGAGATTGAGTTGGAGAGA 185394
271GluGlnAlaIleHisValSerIleArgProGluGlu... 282
185395 ACAATCTCATAGTAGTAAGAGATATTGGATTAAGCCCTGAGATATC 185444
283PheIleLysAspGluSerGlyAspIleGluGlyThrIL 295
185445 TAGAGTCTTCTTTCATTAAGGTTGAA.....GAACCTAGATATCGT 185488
295 eArgAspSerVal.....TyrLeuGlyLeuAsnThrAspTyrP 308
185489 AAGGCCCAAGTTGACATAGTGAAGATCGGTGGGAGAGAAATGATGTC 185538
308 heIleGluThrGly.....PheAlaSerLysIleGlnValSerGlu 321
185539 ATCTAAGTTAGGAGATATTATTATTACTGCGAAGTTC.....CCCGAG 185582
322 GluSerThrPheGluGlu.....AspLeuG 330
185583 GACTCAGGTGTTAAGAGGCCCAAGAGCCTGAGCTATTTCTTTATATGAA 185632
330 nLysGlyAsnArgIleArgLeuArgIleAsnThrGlnLysLeuAsnIleP 347
185633 GAAGGCTCATGTTCAAT.....AAACCATGAAAGGCGTATGCT 185673
347 heSerAlaAspGlySerGlnAsnLeuIleLysGly..... 358
185674 TCTTAAAGTTCTCTTGAGAGAAATCTTGTACTCTTCACTTAACCTCTTC 185723
359 ...ValAsnHisGlyThr 363
185724 CTCTCAACCTTTGGAACG 185741

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embL/NA2001B.DAT: AAS52319
seq_documentation_block:
ID AAS52319 standard; DNA; 1059 BP.
XX
AC AAS52319;


```

XX 13-FEB-0002 (first entry)
DF
XX #E. coli DNA for cellular proliferation protein #41.
DE
XX Antisense; ds: prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
OS Escherichia coli.
XX
XX MO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001MO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD,
PI Yamamoto RT, Xu HH;
PI MPI: 2001-611495/70.
DR P-PSDB; AAU34460.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PS
XX Claim 27; Seq ID No 5956; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential
CC for prokaryotic cellular proliferation, their use in identifying
CC genes, their use in the discovery of novel antibiotics, the e
CC genes themselves and the encoded proteins. The prokaryotes use
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Kili
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis.
CC invention is also useful for the identification of potential
CC for antibiotic development. The antisense nucleic acids can a
CC to identify proteins used in proliferation, to express these
CC and to obtain antibodies capable of binding to the expressed
CC The proteins can be used to screen compounds in rational drug
CC programmes. The antisense nucleic acid sequence is also useful
CC for homologous nucleic acids which are required for cell prol
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC NOTE: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1059 BP; 234 A; 287 C; 310 G; 228 T; 0 other;
SQ

```

[illegible]

318 nvalSerGIuInSerThrPheGluGluAspLeuGlnLysGlyAsnArgI 335
|||||
940TGGCAGCGGAGGAGA 955
335 leArgLeuArgIleAsnThrGlnLysLeu 344
|||
956 TATTATTGACGAGTCAACGCTACGCGTCTG 984
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41224
seq_documentation_block:
ID AAH41224 standard; DNA; 349980 BP.
AC AAH41224;
XX
XX 29-OCT-2001 (first entry)
XX
XX Pyrococcus abyssi genomic fragment #3.
DE
XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
XX Pyrococcus abyssi.
FH
FT Key Location/Qualifiers
FT misc_feature 1..49980
FT /tag= a
FT /note= "this sequence overlaps with the 3' end of
FT AAH41223."
FT misc_feature 300001..349980
FT /tag= b
FT /note= "this sequence overlaps with the 5' end of
FT AAH41225."
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Pliet D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Hellig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX
XX Claim 1; Page 347-443; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223
XX and the 3' end of this sequence overlaps with the 5' end of AAH41225. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110
XX degrees centigrade.
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAB75903-AAH75920 and AAB66436.
XX
XX Sequence, 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;
SQ

alignment_scores:
Quality: 630.00 Length: 320
Ratio: 2.751 Gaps: 5

Percent Similarity: 71.562 Percent Identity: 43.125
alignment_block:
US-09-769-787-162 x AAH41224 ..
Align seg 1/1 to: AAH41224 from: 1 to: 349980
22 leuAsnIleThrIleProLysGlySerLeuPheThrLeuGluAlaLeu 38
: : : : :
154056 GTTGACCTTGAGCTTAAGGAGCGGAGATTATTCATCCCGGCGCCGAG 154105
38 rGlyCysGlyLysThrPheLeuArgMetIleAlaGlyPheAsnSerI 55
: : : : :
154106 CGGTTCGGGCAGACGACGACCTCAGGATATGCGGTTGGAAAGCC 154155
55 leuGluGlyGluPheThrPheAspThrLysIleAsnAspMetGlu 71
: : : : :
154156 CAGACAAAGAAATGCTGTTATTCAGTGTAGGATGATGATCTTGTGCC 154205
72 ProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIlePhe 88
: : : : :
154206 CCTTACTCGAGAAACATAGCGCTTCGTCGAGACTACGCTTGTCC 154255
88 OhIleuThrValArgAspAsnValAlaPheGlyLeuMetClnLysLys 105
: : : : :
154256 ACATATGACGCTTTTCAGAAACGTGGCTTGGCTTGAGTTAGGAAC 154305
105 alProLysGluGluLeuIleGlnIleThrAsnLysTyrLeuGluMet 121
: : : : :
154306 TTCCAGCAAAAGACATACAGAGAGAGGAGGAGGAGGAGGAGGAG 154355
122 GlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyGly 138
: : : : :
154356 GCTTGAAGCGGCTTGAAGATCGCATGCCGAGAACATTAACCGGAGTCA 154405
138 nGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAspProSerVal 155
: : : : :
154406 ACAACAGAGGGTTCCTTGAAGGCTTGGCTTGAAGGAGGAGGAGG 154455
155 eulMetAspGluProLeuSerAsnLeuGluAlaLysLeuArgLeuAsp 171
: : : : :
154456 TCCTCTTAGAGCACTCTTAAGCAATCTAGATGCGCAAGGTTAGAGAGC 154505
172 MetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrTh 188
: : : : :
154506 CTGAGATGAGATAAAAAGATACAGAGAGGTTGGCATTAACGAGAT 154555
188 lTyrValThrHisAspGlnGluGluAlaMetAlaIleSerAspGlnIle 205
: : : : :
154556 GTACGTTAACCCATGACCAAGAGGCCATGCTATTAAGCAGCAATAG 154605
205 laValMetLysAspGlyValIleGlnGlnIleGlyArgProLysGluLeu 221
: : : : :
154606 CGGTCAATGAACTTCGGAAGATTAAGCAGCTTGAAGGCCCTAGACTC 154655
222 TyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgTh 238
: : : : :
154656 TACTAACACCGAGACGAGACTTCCTCAAAATTCCTCGAAGCGGAAA 154705
238 nIleIleProAlaAsnLeuGluLysArgSer..... 248
: : : : :
154706 TTTCG.....CTAAAGCTCCCTTCACGACGCGCTGCATGCT 154743
249aspGlyAlaTyrIleValPhe 255
: : : : :
154744 TGGGAGAGCTCTGCTTCAAGCTCGGGGTGGATGGGCTGTAAAGTTC 154793
256 SerAspGlyTyrAlaLeuArgMetProAlaLeuAspGlnValGluGlu 272
: : : : :
154794 TTTCAGCCTGAAGAGCTTGAAGATA...GAAAGAGAGCGCTTGAAGCTGA 154840
272 nAlaIle.....HisValSerI 278
: : : : :

154841 GGTATGAGTACGAAATACCTCCGGGAGATAGTTTAGCTTAGCG 154890
278 leaIrProGluGluPheIleLysAspGluSer..GlyAspIleGluGlyT 294
154891 TTCAAGTAAAGAAATATATAGCGGAGACTCTTGGCAAGATGAGAG 154940
294 hrIleArGAspSerValTyrLeuGlyLeuAsnThrAspTyrPheIleGlu 310
154941 ATACGAGAAAGATTATATATAAAGTGAAGAGCTTACGCTTATAGAC 154990
311 ThrGlyPhe 313
154991 TCTGGATTTC 154999
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546248
seq_documentation_block:
ID AA546248 standard; DNA: 14674 BP.
XX
AC AA546248;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding novel mar regulated protein (NIMR) #17.
XX
KM mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS Escherichia coli.
XX
PN W0200170776-A2.
XX
PD 27-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07478.
XX
PR 10-MAR-2000; 2000US-188362P.
XX
PA (TUFT) TUFTS COLLEGE.
XX
PI Levy SB, Barbosa TM, Alekshun MN;
XX
DR WPI: 2001-602769/68.
DR P-PSDB: AAU29348.
XX
XX
PT Identifying compounds that modulate a newly identified mar regulated
PT polypeptide activity, useful as antimicrobial compounds, involves
PT contacting the polypeptide with a test compound .
PS Disclosure; Page 266-276; 526pp; English.
XX
XX
CC The invention relates to a method of identifying compounds that modulate
CC a newly identified mar regulated (NIMR) polypeptide activity. The method
CC comprises contacting an NIMR polypeptide with a test compound under
CC interaction conditions, determining the ability of the compound to
CC modulate the activity or expression of the polypeptide, and selecting the
CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC of microbial infections, and in screening for modulators of NIMR
CC expression and activity. These modulators can be used to reduce the
CC infectivity of a microbe on a surface, and the virulence of a microbe in
CC a subject suffering from an infection. AA546232-AA546278 represent
CC Escherichia coli NIMR coding sequences of the invention.
XX
SQ Sequence 14674 BP; 3637 A; 3767 C; 3842 G; 3428 T; 0 other;

alignment_scores:
Quality: 628.00 Length: 358
Ratio: 2.482 Gaps: 6
Percent Similarity: 70.670 Percent Identity: 36.872

alignment_block:
US-09-769-787-162 x AA546248 ..

Align seg 1/1 to: AA546248 from: 1 to: 14674
1 MetSerGluIleLysIleIleLeuAlaLysIleTyrHisAspValPr 17
13495 ATGGGAGAGGTACACTGCATAATGTAACGAAGCCGTGGGAGAGTCT 13544
17 oValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
13545 GGTATCGAAGATATCATATCTGATATCCATGAAGTGAATTCGTGCT 13594
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
13595 TTGTCGACCGCTGTGCTGCGTAATGACATTCCTACTGCGCATGATGCC 13644
51 GlyPheAsnSerIleGluGlyGluPheTyrPheAspThrLysIle 67
13645 GGGCTTGAGACGATCAGCGCGGCGGCTTCATCGGTGAAGAACGGAT 13694
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnT 84
13695 GAATGACACTCCGCCAGCAGAACCGCGCTGTGATGCTGTTCAGTCTT 13744
84 yValIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
13745 ACGCGCTATCCCACTGCTCAGTACAGAAACATGCTATTTGGCTG 13794
101 MetGlnLysLysValProLysGluGluLeuIleGlnGlnThrAsnLysTy 117
13795 AAAGTGGCTGGCGCAAAAAGAGTGATTAACCAACGGTTAACCGGT 13844
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysL 134
13845 GCGCGAAGTGTCTACAATGCGGCATTTGCTGATGCTCAAAACCGAAAGCC 13894
134 euSerGlyGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
13895 TCTCCGCTGTACGCTCAGCGTGCATGCGCATTTGGCGGTGGGCGC 13944
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAla 167
13945 GAGCCAAAGGATTTTGTCTCATGAACCGCTCTCCAACTGATGCTCTC 13994
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
13995 ACTGCGTGTGCAAAATGCTATCGAAATCTCCGCTGCTAAACGCTTG 14044
184 lYleThrThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200
14045 GCCGCACANTGATTTACGTACCCACGATCAGATGCGATGAGCGCTG 14094
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
14095 GCCGCAAAATGCTGTGCTGAGACCGCGTCCGCTGCGCCAGGTGGGA 14144
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234
14145 ACCGCTGGAGCTGTACCACTATCCGCAACCGCTTTGTGCGCGATTTA 14194
234 lGly.....ArgThrAsnIleIleProAlaAsnLeuLysArgSer 248
14195 TCGGTTCCGCCAANAAGATCACTTCTGCGGTAA..... 14229
249 AspGlyAlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAl 265
14230GTGACCGCCACCGC 14243
265 aleuAspGlnValGluGluGlnAlaIleHisValSerIleArgProGluG 282
14244 AATCGATCAAGTGCAGGTGAG...CTGCCGATGCCAAATGCTGACGAAG 14290
282 lPheIleLysAspGluSerGlyAspIleGluGlyThrIleArgAspSer 298
14291 TCTGGCTGCCAGTTGAAGCCGATGCTCAG.....GTGAGAGCAAT 14334

```

299 ValTyrLeuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSe 315
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
14335 ATGTCCCTGGGTATTCGCCGGAACATCTACTGCCGAGTATCGCTGA 14384
315 LysIleGlnValSerGlnGluSerThrPheGluGluAspLeuGlnLysG 332
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
14385 CGTCATCTCTTGAGGTCAGTTCAGGTCGTCGAG.....CAACTCG 14425
332 LysAsnArgIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSer 348
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
14426 GCACAGCAAACTCAATCCATATCCAGATCCCTTCATTCTG..... 14466
349 AlaAspGlySerGlnAsnLeuIle 356
14467 .....CAAAACCTGCTG 14478

seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF94377
seq_documentation_block:
ID   AAF94377 standard; DNA; 984 BP.
XX
AC   AAF94377;
XX
DT   04-JUN-2001 (first entry)
DE   Haemophilus influenzae essential bacterial gene SEQ ID NO:65.
XX
KW   Haemophilus influenzae; essential bacterial gene; identification;
KW   otitis media; meningitis; upper respiratory tract infection;
KW   infection; antimicrobial; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO20011033-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US21176.
XX
PR   04-AUG-1999; 99US-0368382.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Chovan LE, Hessler PE, Reich KA.
XX
DR   WPI: 2001-147511/15.
XX
P-PSDB: AAB88524.
XX
PT   Essential bacterial genes from Haemophilus influenzae and methods for
PT   identifying 'essential' genes that may be potential therapeutic targets
XX
PS   Claim 2; Page 107-109; 185pp; English.
XX
CC   AAF94345 to AAF94409 represent essential bacterial genes from
CC   Haemophilus influenzae, which encode the proteins given in AAB88492 to
CC   AAB88556. The present invention also describes methods for identifying
CC   essential bacterial genes (i.e., those essential to the survival of a
CC   bacterium) using a transposition system. The methods are used to
CC   identify essential genes from bacteria, especially H. influenzae (which
CC   causes otitis media, meningitis and upper respiratory tract infections)
CC   which may be used as targets for potential antimicrobial agents.
CC   AAF94410 to AAF94416 represent PCR primers used in the exemplification
XX
SQ   Sequence 984 BP; 300 A; 160 C; 236 G; 288 T; 0 other;

```

```

alignment_block:
US-09-769-787-162 x AAF94377
Align seg 1/1 to: AAF94377 from: 1 to: 984

8 AsnAlaLysLysIleTyrHisAspValProValIleGluAsnLeuAsnI 24
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
31 AATATCAAAAAGCATTGGTAAAGCGGCTGCATCATCATTAATTAGATT 80
24 ePrrIleProLysGlySerLeuPheThrLeuGlyAlaSerGlyCysG 41
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
81 AACGATCAAAAGTGCGCACATGTACATTTGATGGCCATCAGGCTGTG 130
41 LysThrThrLeuLeuArgMetIleAlaGlyPheAsnSerIleGluGly 57
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
131 GTAAACCACCGTATTACGTTAGTGGCAGAGATTAGAAATTCACACATCA 180
58 GlyGluPheTyrPheAspAspThrLysIleAsnAsnMetLupProSerL 74
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
181 GGTCAAAATATTATTGATGGCGAAGATGTAAACAAATCCTATTTCAGAA 230
74 sArgAsnIleGlyMetValPheGlnAsnTyrAlaIlePheProHisLeu 91
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
231 TCGAGATATTGTAATGTTTCCATATCTTACGCGCTTTTCCCGCATATGA 280
91 hrValArgAspAsnValAlaPheGlyLeuMetGlnLysLysValProLys 107
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
281 GCATTTGGCGCATTAACGTGCGCTACGCGTTAAATAATGCAGACATTGGCATA 330
108 GlnGluLeuIleGlnGlnThrAsnLysTyrLeuGlnLeuMetGlnIleAl 124
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
331 GAAGAACGCGCTCAGCGCTAAAGAGAGCTTGAATTAAGTGAATTAGC 380
124 agLntYrAlaAspArgLysProAspLysLeuSerGlyGlnGlnGlnAla 141
   | :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
381 GGGCTTTGAAGATCGTTTCTGCGATCAATTTCTGGGGCACAACACAC 430
431 GTGCGCATTTGGCTCGTGGTGGTATTAACCAACAAAGCTCATATGTTT 480
158 AspGluProLeuSerAsnLeuGlnAlaLysLeuArgLysPheArgGly 174
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
481 GATGAACCATTAAGTAATTAAGATGCAAACTACGTCGTTCTATGCTGA 530
174 naIaIleArgGluIleGlnHisGluValGlyIleThrThrValTyrVal 191
   : |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
531 AAAAATCCGTAATTCGACACACGTTAGCATTTACCTCCCTTATGTGA 580
191 hrHisAspGlnGluGluAlaMetAlaIleSerAspGlnIleAlaValMet 207
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
581 CACACGATCAACAGACGATTTGGGTATCTGATGAAGATGATGTATG 630
208 LysAspGlyValIleGlnGlnIleGlyArgProLysGluLeuTyrHisL 224
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
631 AATAAAGGTAATTTATGCAAAAAGCGCGCGAAGAGCTTATATCTCG 660
224 sProAlaAsnGluPheValAlaThrPheIleGlyArgThrAsnIleIle 241
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
681 ACCAAATTCCTGTTTGGCTAATCTTATGCGGCAATCAGATATTTCG 730
241 roAlaAsnLeuGluLys..... 246
731 ATGGAATAATTAGAAATATGCGGTGCGGATATTAATGTTACTGTGCGCT 780
247 ArgSerAspGlyAlaTyrIleValPheSerAspGlyTyrAlaLeuArg 263
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
781 TTAAAGATGCTGCACAGTTTAATTAACCGATGCGCAAGTTTA..... 825
263 tProAlaLeuAspGlnValGlnGluGlnAlaIleHisValSerIleArg 280
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
826 .....GTGGGTATTCGCC 838

```

```

alignment_scores:
Quality: 627.50      length: 326
Ratio: 2.637         Gaps: 4
Percent Similarity: 73.006  Percent Identity: 41.104

```


[illegible]


```

470 CGTCCGGGTGTGGAGACGACGACGATTGGCATGATCGCGGATTCGAG 519
54 SerIleGluGlyGlyPheTyPheAspThrIleAsnAsnAsn 70
520 ACCCGGCTAAGGGGATCCGCTCGAAGGCCGACGCTGCGAGGAC 569
70 TgluProSerIleArgAsnIleGlyMetValPheGlnAsnTyAlaIleP 87
570 CCGACCCAAAGCGCAACGTCACAGGTGTCACGACTACGCGCTGT 619
87 heProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
620 TCCCGACATGAGGTGTGGACACGTCGCTACGCGCCGCGACGCAAG 669
104 LysValProLysGluGluLeuIleGlnGlnThrAsnLysTyLeuGlu 120
670 AAACTCGGCAAGCGGAGTCCGCAAGCGGCTGACAGAGCTGCTGGAGAT 719
120 uMetGlnIleAlaGlnTyAlaAspArgLysProAspLysLeuSerGly 137
720 CGTCCGGCTACCGAATTGTCGCGAGCGACGCCCGCCAGCTGTCGCGG 769
137 LysGlnGlnArgValThrLeuAlaLysAlaLeuAlaValAsnProSer 153
770 GCGAGCGACGCGGCGGTGGCTGGCGGCGGACTGTGAACCTACCCAGC 819
154 ValLeuLeuMetAspGluProLeuSerAsnLeuGlnValLysLeuArg 170
820 GCGCTGTGCTGATGAACCGCTCGAGCGCTGACCTGAAGCTGCGGCA 869
170 uAspMetArgGlnAlaIleArgGluIleGlnIleGlnValGlyTleThr 187
870 CGCATCATGCTGAGCTCAAGCGCATCAAGCGGAGGTGCGGATCACT 919
187 hValTyValThrHisAspGlnGluValMetAlaIleSerAspGln 203
920 TCATCTACGACCGACGACGACGAGGAGGCGCTCACGATGACGACGC 969
204 IleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 220
970 ATGCGGCTGATGAACGCGGCGGACGCTGAACAGATCGGACCGGACCA 1019
220 uLeuTyHisLysProAlaAsnGluPheValAlaThrPheIleGlyArg 237
1020 GATCTACGACCGTCCGCGGAGGCTGTGTCGCGCAGCTTATCGACAG 1069
237 hTrsAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTy 253
1070 CCAACCTCTGGCGGCGGCTGACCGCGGCTCAACCGCGATTTACGTC 1119
254 ValPheSerAsp...GlyTyAlaLeuArgMetProAlaLeuAspGln 269
1120 GAGATCGACCTTCGCGCTGACGCTGAAGCGACCGCGGCGGACGACCA 1169
269 LgluGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
1170 GATCGAGCGCGGCGGACGACCACTGATGTCGTCGGAACGATCC 1219
284 .....IleLysAsp...GluSerGlyAsp.....Ile 291
1220 GGGTCAACCGCGGCTCCGAGACGCGCGGCTGACGTCGCTGCGTG 1269
292 GluGlyThrIleArgAspSerValTyLeuGly 302
1270 CGTCCGCTACCGGCTGACGCTTCCAAAGT 1302

```

seq_name: /cgn2.6/plodata/1/lna/5B.COMB.seq:US-08-997-362-88

seq_documentation_block:
 : Sequence 88, Application US/08997362
 : Patent No. 5985287
 : GENERAL INFORMATION:
 : APPLICANT: Tan, Paul

```

? APPLICANT: Miyama, Jun
? APPLICANT: Visser, Elizabeth
? APPLICANT: Skinner, Margot
? APPLICANT: Scott, Linda
? APPLICANT: Prestidge, Ross
? TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
? NUMBER OF SEQUENCES: 194
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Law Offices of Ann W. Speckman
? STREET: 2601 Elliott Avenue, Suite 4185
? CITY: Seattle
? STATE: WA
? COUNTRY: USA
? ZIP: 98121
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/997,362
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
? FILING DATE: June 12, 1997
? APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
? FILING DATE: August 29, 1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Sleath, Janet
? REGISTRATION NUMBER: 37,007
? REFERENCE/DOCKET NUMBER: 11000.1002c2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 206-269-0565
? TELEFAX: 206-269-0563
? TELEX:
? INFORMATION FOR SEQ ID NO.: 88:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1518 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? US-08-997-362-88

```

alignment_scores:
 Quality: 610.00 Length: 311
 Ratio: 2.652 Gaps: 5
 Percent Similarity: 73.955 Percent Identity: 42.444

alignment_block:

US-09-769-787-162 x US-08-997-362-88 ..

Align seq 1/1 to: US-08-997-362-88 from: 1 to: 1518

```

4 IleLysIleIleAsnAlaLysIleTyHisAspValProValIleG 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 ATCGAGATCGACCATGTCACGACGCTTGGCGGCTGACGCTGCGC 419
20 uAsnLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGlu 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
420 GAGCGGAGATCTTCATCGCGCGCGGAGGAGTTCTTCATGCTCGC 469
37 lAserGlyCysGlyLysThrIleLeuArgMetIleAlaGlyPheAsn 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
470 CGTCCGGGTGTGGAGACGACGACGTTGGCGCATGATCGCGGATTC 519
54 SerIleGluGlyGlyPheTyPheAspThrIleAsnAsnAsn 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
520 ACCCGGCTAAGGGGATCCGCTCGAAGGCCGACGCTGCGAGGAC 569

```

```

70  tGluProSerLysArgAsnIleGlyMetValPheGlnAsnThrValIleLeu 87
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
4570  CCACCCAAACAAACGGCAAGCTCAACACGGGTGTTCCACACTACCGCGCTG 619
87  heProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
620  TCCCGGACATGAGAGGTCGTGGGACACTCCGCTACGGCCCGCGAGCAG 669
104  LysValProLysGluGluLeuIleGlnGlnThrAsnLysTyLeuGluLe 120
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
670  AAACCTGGCAAAAGCAGAGTCCCAAGACGGCGTGCAGCAGCTGTGAGAT 719
120  uMetGlnIleAlaGlnIleThrAlaAspArgLysProAspLysLeuSerGly 137
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
720  CGTCCGGGTGACCGAATTTGCGCAGCGACGGCCCGCCACGTTGTCGGCG 769
137  IyGlnGlnGlnArgValAlaThrLeuAlaCysAlaLeuAlaValAlaProSer 153
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
770  GGCAGCAGCAGCGGGGTGGCTTGGCCCGGGGACACTGTGAACTACCCAGC 819
154  ValLeuLeuMetAspGluProLeuSerAsnLeuGlnAlaLysLeuArgLe 170
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
820  GGGCTCTGCTCTCATGAACCGCTCGGAGCGCTGCAGCTGAAGCTGCGCCA 869
170  uAspMetArgGlnAlaIleArgGluIleGlnHisGlnValGlyLeuThrT 187
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
870  CGTCAATGCGAGTTCGACATCAAGGCGATCAAGCGGAGGTGGGTATTCAGT 919
187  hrValTyAlaThrHisAspGlnGluGlnAlaMetAlaIleSerAspGln 203
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
920  TCATCTACGTGACCCACGACCAAGAGAGCGCTCACGATGATGACCCG 969
204  IleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLysGlu 220
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
970  ATCGCGGTGATGAAGCGCGGCAACGTGCAGAACATGTGGCAGCCCGACCGA 1019
220  uLeuTyHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1020  GATCTACGACCGGCTCCGCGAGCGGTGTGTGCGCAGCTTCATCGCAGAG 1069
237  hrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyLe 253
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1070  CCAACTCTGTGGCGGGCGGTGCAACGGCGCTCAACCGCGATTCAGTC 1119
254  ValPheSerAsp...GlyTyAlaLeuArgMetProAlaLeuAspGlnVal 269
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1120  GAGATGAGAGTTCGCGCTCGACGCTGAAGGACAGCCCGGCGAGACAC 1169
269  IglGluGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1170  GATCGAGCGCCGCGGGCGACCCACCCCTGATGATGTCGTCGGAAGCATCC 1219
284  .....IleLysAsp...GluSerGlyAsp.....Ile 291
1220  GGCTACCCCGGGCTCCACAGAGAGCGCGGACGGGTGACGTGCGCTGCGTG 1269
292  GluGlyThrIleArgAspSerValTyLeuGly 302
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1270  CGTGGCAGCGTCAACGACCTGACCTTCAGAGT 1302
seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-873-970-88
seq_documentation_block:
: Sequence 88, Application US/08873970
: Patent No. 6001361
: GENERAL INFORMATION:
: APPLICANT: Tan, Paul
: APPLICANT: Hayama, Jun
: APPLICANT: Visser, Elizabeth
: APPLICANT: Skinner, Margot
: APPLICANT: Scott, Linda
: APPLICANT: Prestigde, Ross
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR

```

```

1 TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
2
3 NUMBER OF SEQUENCES: 106
4
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: Law Offices of Ann W. Spekman
7 STREET: 2601 Elliott Avenue, Suite 4185
8 CITY: Seattle
9
10 STATE: WA
11
12 COUNTRY: USA
13
14 ZIP: 98121
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Diskette
18
19 COMPUTER: IBM Compatible
20
21 OPERATING SYSTEM: DOS
22
23 SOFTWARE: FastSeq for Windows Version 2.0
24
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/08/873,970
27
28 FILING DATE:
29
30 CLASSIFICATION: 435
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: 08/705,347
34
35 FILING DATE: 29-AUG-1996
36
37 ATTORNEY/AGENT INFORMATION:
38 NAME: Sleath, Janet
39
40 REGISTRATION NUMBER: 37,007
41
42 REFERENCE/DOCKET NUMBER: 11000.1002c1
43
44 TELECOMMUNICATION INFORMATION:
45 TELEPHONE: 206-269-0565
46
47 TELEFAX: 206-269-0563
48
49 TELEX:
50
51 INFORMATION FOR SEQ ID NO: 88:
52
53 SEQUENCE CHARACTERISTICS:
54 LENGTH: 1518 base pairs
55 TYPE: nucleic acid
56 STRANDEDNESS: single
57
58 TOPOLOGY: linear
59
60 MOLECULE TYPE: Genomic DNA
61
62 US-08-873-970-88

```

```

alignment_scores:
  quality: 610.00      Length: 311
  Ratio: 2.652        Gaps: 5
  Percent Similarity: 73.955    Percent Identity: 42.444

Alignment block:
US-09-769-787-162 x US-08-873-970-88  ..

Align seg 1/1  to: US-08-873-970-88  from: 1  to: 1518

4  TlelylllelleasAlAlLyslysllelyrhtlsAspAlProvallelgl 20
|||||  :::::  |||  ::  |||  :::::
370 ATCGAGATCGACCAATGCACGAGCGCTTGGCGACATACCTGGCGCTGC 419
20 uAsnleuAsnleThrllleProlylGserleuPheThrleuLeuAla 37
::  :::::  ||||  |||  :::::  ||||  |||  :::::  ||||  |||
420 GGACGAGACCTTCACATCGGGCCCGGGAGATTCTCTCCATGCTCGGCC 469
37 lAserylCYsGlylYsThrThrleuLeuAurgetlleAlaglyPheasn 53
|||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
470 CGTCCGGGTGTGGGAAGACGACACCGTTCGCCATGATCGCGGATTCGAG 519
54 SerlleglUGlyGluPheThyrPheAspAspThrlyslleAsnAsme 70
:::  |||  :::  :::  :::  :::  :::  :::  :::  :::
520 ACCCGCATGAAAGGGGCGCATCGCTCGCAAGCGCCGACGTGTGAGAC 563
70 tGluProSerlyAsrAsnlelleGlyMetValPheGlnAsnThrAlalle 87
|||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
570 CCACCCACAACGACGCAACGTCAACACGGGTTCACAGACTACACGCTGT 619
87 heProHlsleuThrAlArgrAspAsnValAlaPheGlyLeuMeGlnLys 103
|||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
620 TCCCGCATGTCAGCGGTCTGGGACACAGTCGCGTACGGGCCCGCAGGAG 669

```



```
720 CGTGGCGCTGACCGAATTTGCCAGCGAGCCGCCCGCAGCTGCCGCG 769
437 IYGLINGINLARGVALTHRLAVALCYALALALALALALALALALAL 153
770 GGCAGAGACGCGGGTGGGTTGGCCGGGCGACTGCTGAATCACTCCAC 819
154 VALLEULEMETASPLUPROLEUSERANLEUGLUALALALALALAL 170
820 GCGGCTGCTCTCGATGAGACCGCTCGAGCGCTCGACCTGACCTGCG 869
170 uaspmetarGINAlleIlearglulIleGlnhISgluValIglYleThT 187
870 CCGTATGCGAGTTCGAGCTCAAGCCGATCCAGCGGAGTCCGGATCAC 919
187 hValTYrValThrhISAspGlnGluGluAlaMetAlaIleSerAspGln 203
920 TCATCTACGTGACCCAGACGAGAGAGCGCTCACGATGAGTACCCGC 969
204 ILeAlaValMeLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
970 ATCGGCGGTATGACGCGCGCAACGTGCAACAGATCGCAGCCGACCG 1019
220 uLeuTYrHISLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
1020 GATCTACGACCGCTCCCGCAGCGGTGTTCTCGCCAGCTTCATCGCAG 1069
237 hAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrlle 253
1070 CCAACCTCTGGGCGGCGGCTCACCGCGCGCTCCCAACCGCGATTCAGT 1119
254 ValPheSerAsp...GlyTYrAlaLeuArgMetProAlaLeuAspGlnVa 269
1120 GAGATCGACGCTCTCGCTCGACGCTGAGAGCAGCGCGCGGAGACCA 1169
269 IgluGlnGlnAlaIleHISValSer.....IleArgProGluGluPhe. 283
1170 GATCGAGCGCGGCGGCGGACCACTGATGGCGTCCGGAACGATCC 1219
284 .....IleLysAsp...GluSerGlyAsp.....Ile 291
1220 GGGTACACCCGGGCTCCAGAGCGCGCGAGCTGACGTGCGCTGCGT 1269
292 GluGlyThrIleArgAspSerValTYrLeuGly 302
1270 CGTGCCACCGCTCACCGACCTGACCTTCCAAGGT 1302

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-324-542-88

seq_documentation_block:
: Sequence 88, Application US/09324542
: Patent No. 6328978
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Tan, Paul L.J.
: TITLE OF INVENTION: Methods and Compounds for the Treatment
: TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
: FILE REFERENCE: 11000.1007c1
: CURRENT APPLICATION NUMBER: US/09/324,542
: EARLIER FILING DATE: 1999-06-02
: EARLIER APPLICATION NUMBER: US 08/997,080
: NUMBER OF SEQ ID NOS: 194
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 88
: LENGTH: 1518
: TYPE: DNA
: ORGANISM: Mycobacterium vaccae
US-09-324-542-88

alignment_scores:
Quality: 610.00 Length: 311
```

```
Ratio: 2.652 Gaps: 5
Percent Similarity: 73.955 Percent Identity: 42.444
alignment_block:
US-09-769-787-162 x US-09-324-542-88 ..
Align seg 1/1 to: US-09-324-542-88 from: 1 to: 1518
4 IleLysIleIleAsnAlaLysLysIleTYrHISAspValProValIleG 20
370 ATCGAGATGACCATGTCACGAGAGCGCTTCGCGACTACCTGCGCTGC 419
20 uasLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGlyA 37
420 GAGACGAGACTTCTTCATCGCGCGCGGAGTTCCTTCATGCTCGGCC 469
37 IASerGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsn 53
470 CGTCCGCGGTGGGAAGACGACACGTCGTCGATGATCCGGGATTCGAG 519
54 SerIleGluGlyGlyLuhPheTYrPheAspAspThrLysIleAsnSme 70
520 ACCCGGACTGGAAGGGGCGATCCGCTCGAAGCGCGGACGTGTCGAGAC 569
70 tGluProSerLysArgAsnIleGlyMetValPheGlnAsnTYrAlaIleP 87
570 CCCACCCACAAAGCGACAGCTCACACGSGTGTCCAGCATACGCGCTGT 619
87 hEProHISLeuThrValaLargAspAsnValAlaPheGlyLeuMetGlnLys 103
620 TCCCGCATGACGCGTGGGACAAACGTGCGAGACGCGCGCGGACAGAG 669
104 LysValProLysGluGluLeuIleGlnGlnThrAsnLysTYrLeuGlu 120
670 AAACCTGGCAAGGCGAGGTCCGCAAGCGCGTCGAGACCTGCTCGAGAT 719
120 uMetGlnIleAlaGlnTYrAlaAspArgLysProAspLysLeuSerGlyG 137
720 CGTCCGCGCTGACCGAATTTGCCGAGCGAGCGCGCCGCGAGCTGTCCGG 769
137 IYGLINGINLARGVALTHRLAVALCYALALALALALALALALALAL 153
770 GGCAGAGACGCGGGTGGCGGTGGCGCGGCACTGCTGAATCACTCCAC 819
154 VALLEULEMETASPLUPROLEUSERANLEUGLUALALALALALAL 170
820 GCGCTGCTGCTGATGAAACCGCTCGAGCGCTGACCTGACCTGCGCA 869
170 uaspmetarGINAlleIlearglulIleGlnhISgluValIglYleThT 187
870 CCGTATGCGAGTTCGAGCTCAAGCGCATCCAGCGGAGTCCGGATCAC 919
187 hValTYrValThrhISAspGlnGluGluAlaMetAlaIleSerAspGln 203
920 TCATCTACGTGACCCAGACGAGAGAGCGGCTCACGATGAGTACCGC 969
204 ILeAlaValMeLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
970 ATCGGCGGTATGACCGCGCAACGTGCAACAGATCGCAGCCGACCGCA 1019
220 uLeuTYrHISLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
1020 GATCTACGACCGTCCGCGAGCGGTGTTCTGCTGCGCAGCTTCATGCGCA 1069
237 hAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrlle 253
1070 CCAACCTCTGGGCGGCGGTTGACCGCGCGTCCCAACCGCGATTCAGT 1119
254 ValPheSerAsp...GlyTYrAlaLeuArgMetProAlaLeuAspGlnVa 269
1120 GAGATGACGCTTCTCGCTCGAGCTGGAAGGACCGCGCGGAGACCA 1169
269 IgluGlnGlnAlaIleHISValSer.....IleArgProGluGluPhe. 283
```



```

seq_name: /cgn2.6/ptodata/1/lna/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

```

```

alignment_scores:
Quality: 583.50      Length: 367
Ratio: 2.372         Gaps: 6
Percent Similarity: 67.030      Percent Identity: 35.695

```

alignment_block:

US-09-769-787-162 x US-09-103-840A-1/rev ..

Align seq 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529

```

1 Metergultlelylleleasnlalsylslyr.....HISAS 15
|||||.....:|||||:|||||
2284794ATGGCTTCGGAGCTTTTGACAGGCAACCGCGCTATCCCGCACCGA 2284745
15 pvalprovalilleglusnleuasnllethrleprolysglyserleup 32
| |||||.....:|||||:|||||
2284744ACGACCGCGCTGATCGCTCGACACCTATCGTCGGCGCATGGCGATTG 2284695
32 herthrleuengllyalsersglycylslythrthrleuLeuArymet 48
:|||||:||||| ||||| ||||| ||||| |||||
2284694ATTGCTGCTGGCGCGCTCGGATGTGGCAGACGACGTCGTACGATG 2284645
49 llelaaglypheasenseriellegluglyluphetryrphesaspaspth 65
:|||||:|||||:|||||:|||||:|||||:|||||
2284644GTGGCTGGCTGGAGACGCTGACTGTGGCGCTATCCGATCGGCGACCG 2284595
65 rlyslleasnasmetgluproserylsargasnilleglymetvalphag 82
:|||||:|||||:|||||:|||||:|||||:|||||
2284594ACGACGTCACCGAGTCGATCCCAAGATCGTGAATGCGCCATGGTGTCC 2284545
82 lnasnlyrAlaIlePheProHISLeuThrValArGspasnValAlaPhe 98
|||||:|||||:|||||:|||||:|||||:|||||
2284544AAGACTACGCTTACCGCCGACATGACGGCGGCGCAAGACATGGCGTCT 2284495
99 GlyleuemetGlnLysValProlysGlnLueulleGlnInThrAs 115
:|||||:|||||:|||||:|||||:|||||:|||||
2284494GCGGTGAAGTCGCCAAGATCGGCAAGGATCCGCGCGCGGCT 2284445
115 nlytyrleuengluleuemetGlnlelaGlnlyrAlaAspArgLysPro 132
:|||||:|||||:|||||:|||||:|||||:|||||
2284444ATGCCGCGCAAAATTCCTGATCTGCATCTTATCTGATCGCAAGCCGA 2284395
132 splysleuSerglyGlyngInGlnArGValThrleuAlaCysAlaLeu 148
* ||||| ||||| ||||| ||||| ||||| |||||
2284394AAGATCTCTCGGCGCGCAACGCGGATGGGCGATGGCGATC 2284345
149 AlAvalasnProSerValleuLeuMetaspGluProleuSerAsnLeuG 165
:|||||:|||||:|||||:|||||:|||||:|||||

```

```

2284344GTGGCGGCCCCACAGGATTCCTGATGAGACGACCGCTGCCAAATCTTGA 2284295
165 uAlaLysLeuArgLeuAspMetArgInAlaIleArgGluIleGlnHis 182
||||| ||||| ||||| ||||| ||||| |||||
2284294AGCCAAATTCGGCGGCGCAACCGCCATTCAGATCGCGCGTTACAGCGC 2284245
182 luValGlyIleThrValTyValThrHisAspGlnGluAlaMet 198
:||||| ||||| ||||| ||||| ||||| |||||
2284244AACTGCTACGACCCAGCTGATGTCACTACGACGAGTCGAGGCGATG 2284195
199 AlAlleSeraspGlnleAlaValMetLysAspLysAlaIleGlnInI 215
:|||||:|||||:|||||:|||||:|||||:|||||
2284194ACGATGGGTACCGCGTGGCTGTCTGCTACCGCTGCTGACACAGTG 2284145
215 eGlyArgProlysgLueuTyThrHisLysProAlaAspGluPheValAla 232
:|||||:|||||:|||||:|||||:|||||:|||||
2284144ATGCTTCGCTGAGAGCTTACCGCCAAACCGGCGACAGTGTGTCGCGG 2284095
232 hrPheIleGlyArgThrAsnIleleProAla..AsnLeuGluLysArg 247
||||| ||||| ||||| |||||
2284094AGGTCATCGCTTCC.....CCGCGATGAACCTGTTCAGGCTT 2284057
248 SerAspGlyAlaTyrlleValPheSerAspGlyTyrlaLeuArgMetPr 264
||| :||| |||||
2284056TCCATCGCCGATTCACCGGTCTCACTGGGTGATGGCAGATCTGTGCTG 2284007
264 cAlaLeuAspGlnValGlnGlnAlaIleHisValSerIleArgProG 281
| |||||.....:|||||:|||||
2284006GCGTGGCGGTGCTGCGTACCGGCGGAGGTGATATGCGGTGTCGCCCG 2283957
281 lugiPheIleLysAspGluSerglyAspIleGluGlyThrIleArgAsp 297
|||||:|||||:|||||:|||||:|||||
2283956AACATTTGGAGCTGGCGGCGCGCGC..... 2283931
298 SerValTyrlleuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAl 314
:|||||:|||||:|||||:|||||:|||||
2283930.....ATCGATGAGACGTGCACATGCTGCAGAACTGTGAGCGCGA 2283890
314 aSerLysIleGln.....ValserGluGluSerThrPheGluGluA 328
:|||||:|||||:|||||:|||||:|||||
2283889GCGCTATGTATGCCGAATCGTGTGCGGCGGCTGCGAAATGGACAGT 2283840
328 splEu.....GlnLysGlyAsnArg 334
:|||||:|||||:|||||:|||||:|||||
2283839CAATCGCTCGATGAGCGCGCGCGCGCGCGCGCGGAGGAGTACGTC 2283790
335 lleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAspG 351
:|||||:|||||:|||||:|||||:|||||
2283789GTGGGCTATGTCCACACCGGCGGACACCTGCATCTTTCGCGTGCAGG 2283740
351 y 351
2283739G 2283739

```

seq_name: /cgn2.6/ptodata/1/lna/6B_COMB.seq:US-09-103-840A-2

```

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765

```



```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
  Quality: 583.50      Length: 367
  Ratio: 2.372        Gaps: 6
  Percent Similarity: 67.030  Percent Identity: 35.695

```

alignment block:

US-09-769-787-162 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

1 MetSerGluLeuLysIleIleAsnAlaLysLysIleTyr.....HisAs 15
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2287093ATGGCTTGAGTTGAGCAGGCAACCGCGCTATCCCGGACGGA 2287044
15 pValProValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuP 32
| |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2287043CCGACGCGCGCTGATGCGCTCGACCTGATCGCGCGATGGCGAGTTG 2286994
32 hethrLeuLeuGlyAlaSerGlyCysGlyThrThrLeuLeuArgMet 48
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286993TTGTCCGTGGGGCCGTCGCGATGTGGCAAGACGACGTGCTACGGATG 2286944
49 IleAlaGlyPheAsnSerIleGluGlyGluPheThrPheAspPhe 65
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286943GTGGCTGCTTGGAGACGCTGAGCTGGCGGTATCCGATCGCGGAGCG 2286894
65 rLysIleAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheG 82
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286893ACGACGTCACGAGGTGATCCCAAGATCGTATGTCGATGCTGTTC 2286844
82 lnsnTyrrAlaIlePheProHisLeuThrValAlaArgAspAsnValAlaPhe 98
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286843AGAACTACGCTCTACCCGACATGACGGGGGCGGCAACATGGCTTC 2286794
99 GlyLeuMetGlnLysValProLysGluGluLeuIleGlnGlnThrAs 115
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286793GCGTTGAGTCCCAAGATCGGCAAGCGGAGATCCGCGGCGGTGCT 2286744
115 nLysTyrrLeuGluLeuMetGlnIleAlaGlnTyrrAlaAspArgLysPro 132
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286743TTCGCGACGAATGCTGTATCTGCAATCTTATCTGATCGCAAGCGGA 2286694
132 sPlyLysLeuSerGlyGlyGlnGlnIleArgValThrIleAlaCysAlaLeu 148
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286693AAGATCTCTCCGCGCGCAACGCGGATGGGCGGATGGCGTGCATC 2286644
149 AlaValAsnProSerValIleLeuMetAspGluProLeuSerAsnLeuG 165
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286643GTGGGGGCGCCACAGGTATTCGTGATGACGACCGCTGTCCTCAATCTTGA 2286594
165 uAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisG 182
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286593GCGCAAACTTCGCGCGCAACCGCATCGATCGCGCGCTTACACAGGC 2286544
182 luvAlaGlyIleThrThrValTyrrValThrHisAspGlnGluAlaMet 198
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286543AACAGGTGACGACCGCTGTATGTCACTCACGACCGAGTGGAGGCATG 2286494
199 AlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnI 215
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286493ACGATGGGTGACCGCTCGCGGTCTGTGACGCTGTCTGCGCAACGCTG 2286444
215 eGlyArgProLysGluLeuTyrrHisLysProAlaAsnGluPheValAla 232
::: |||:||||:||||:||||:||||:||||: |||:||||:

```

```

2286443TGCCTCGCCTCGAGAGCTTACCGCAACCGCGCAACGTGTTCGCGG 2286394
232 hrPheIleGlyArgThrAsnIleIleProAla..AsnLeuGluLysArg 247
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286393GCTTCATCGGTTC.....CCGCGATGAACTGTTTCAGGCTT 2286356
248 SerAspGlyAlaTyrrIleValPheSerAspGlyTyrrAlaLeuArgMetP 264
||| ::| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2286355TCCATCGCCGATTCACAGGTGTCTACGTGATGGAGATCTCTGCTGCC 2286306
264 oAlaLeuAspGlnValGluGluAlaIleHisValSerIleArgProg 281
| |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286305GCGTGGCTGTGATGCGGACCGGAGTCATATCGGTGTGCGCCCG 2286256
281 luvGluPheIleLysAspGluSerGlyAspIleGluGlyThrIleArgAsp 297
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286255AACATTGGAGCTGGCGCGCGCGC..... 2286230
298 SerValTyrrLeuGlyLeuAsnThrAspTyrrPheIleGluThrGlyPheAl 314
2286229.....ATCAGATGACGACGTGACATGTGTGCAACACTTGGACGCGA 2286189
314 aSerLysIleGln.....ValSerGluGluSerThrPheGluGluA 328
::: |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286188CGCTACTTGTATGCGCGAATCGTGTGCGCGCGCTGCGAATGACAGAT 2286139
328 sPheLeu.....GlnLysGlyAsnArg 334
::: |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286138CAATCGTGTGATGAGTGAGCGCGCGCGCGCGCGCGGATGATGTCG 2286089
335 IleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAspG 351
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
2286088GTGCGGTATGTCCACACCGGAGACACCTGCTTCTCCGCTCGACGG 2286039
351 y 351
2286038g 2286038
seq_name: /cgn2_6/plodata/1/lna/6B_COMB.seq:US-08-858-207A-52
seq_documentation_block:
; Sequence 52, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858.207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimm1, Edward R
; REGISTRATION NUMBER: 38, 891

```

```
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1828 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-52

alignment_scores:
Quality: 442.00      Length: 90
Ratio: 4.966         Gaps: 0
Percent Similarity: 98.889   Percent Identity: 96.667

alignment_block:
US-09-769-787-162 x US-08-858-207A-52 ..

Align seg 1/1 to: US-08-858-207A-52 from: 1 to: 1828

274 ILEHISVALSERILEARGPROGLUGLUPHEILELYSASPGLUSERGLYAS 290
|||||
3 ATTCATGTGAGACATTCGTCGCCAGAGTTTATCAAGATGATCGAGGA 52
|||||
290 PILEGLUGLUTHRILEARGASPERVALTYRILEUGLYLEUASNTRHAPT 307
|||||
53 TATTAGAGACTATTAGCGATGCGCTCTATCTTGAGCTAAATACGAGGT 102
|||||
307 YRPHLEGLUTHRILEARGPHEALASERLYSILEGLINVALSERGLUGLUSER 323
|||||
103 ATTTCATTGAGACAGGTTTGGCTCAAAAATTCAAGTTAGTCAAGAATCA 152
|||||
324 THRPHLEGLUASPGLUGLINSGLYSANARGILEARGLEUARGILEAS 340
|||||
153 ACTTTTGAAGAGAACTACAAAAGCAATCGTATTCGTCTAGCAATCAA 202
|||||
340 NTHRGINSLEUASNILEPHESERIALASPGLYSERGLINASNLEULEIL 357
|||||
203 TACGCAAAATTTAAACATCTTTCTGCGAGATGTTCCCAAAACGTGATTA 252
|||||
357 YSGLYVALASNHSGLYTHR 363
|||||
253 AAGGAGTCACCATGGAACG 272
|||||

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-578-158-1

seq_documentation_block:
; Sequence 1, Application US/08578158
; Patent No. 5830690
; GENERAL INFORMATION:
; APPLICANT: Govrishankar, Jayaraman
; APPLICANT: Bhandari, Poonam
; APPLICANT: Rajkumari, Kaveti
; TITLE OF INVENTION: USE OF DNA ENCODING A PROMOTER OF
; TITLE OF INVENTION: PROMOTERS ALONG WITH CIS REGULATORY ELEMENTS AND
; TITLE OF INVENTION: A NOVEL PROCESS FOR PRODUCING POLYPEPTIDES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578, 158
FILING DATE: 29-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 07064/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-578-158-1

alignment_scores:
Quality: 418.00      Length: 227
Ratio: 2.580         Gaps: 1
Percent Similarity: 71.366   Percent Identity: 37.004

alignment_block:
US-09-769-787-162 x US-08-578-158-1 ..

Align seg 1/1 to: US-08-578-158-1 from: 1 to: 1260

19 ILEGLASNLEUASNILETHRILEPROLYSGLYSERLEUPHERLEULE 35
|||||
452 GTAAAGACGCCAGCTCTGCCATTCAGAGACGCCGAGATTTTCATCAT 501
|||||
35 UGLYALASERGLYCYSGLYLYSTHRLEULEUARGMETILEALGLY 52
|||||
502 GGGATTATTCGCGCTGCTGTAATCCACAAATGATGCGCTCTCAATGCC 551
|||||
52 HEASNSERILEGLUGLUGLUPHERLYRPHASPTHRLYSILEASN 68
|||||
552 TGATTGAACCCACCCGCCGCAAGTGTGATTGATGTTGATGATATGCC 601
|||||
69 ASNMT.....GLUPROSERLYSARGASNILEGLYME 79
|||||
602 AAATATCCGACGCCGACCTCCGTGAGTGGCGCAAAAAGATTTCGAT 651
|||||
79 TVALPHGLINASNTRYALAILPHEPROHISLEUTHRYVALARGASPASN 96
|||||
652 GGTCTTCAGTCTTGGCTTAATGCCCATATGACCGCTGCGACATTA 701
|||||
96 ALALAPHEGLYLEUMETINLYSVALPROLYSGLUGLULEILEIN 112
|||||
702 CTGCGTTCGTATGGAATTGGCCGGAATTAATGCCGAAGACGCCGGAA 751
|||||
113 GLNTPRASNLYTYRLEUGLULEUMETINILEALAGINTYRALASPAR 129
|||||
752 AAAGCCCTTGATGCATCGCGACAGTGGCGCTGGAATAATTTCGCCAC 801
|||||
129 GLYSPROASPYSLEUSERGLYGLYGLINGLNGLNARGVALTHRYLEUAL 146
|||||
802 CTACCCGGATGACCTCTGCGGGATCGTCAACAGTGTGGATTAAGCC 851
|||||
146 YSALALEUALVALASNPROSERVALLEULEUMETASPGLUPTOLEUSER 162
|||||
852 GCGCGTTAGCATTAATCCGATATTAATTAAGAGAACCCCTTCG 901
|||||
163 ASNLEUGLIALALYSLEUARGLEUASPHEMARGINALALARGLUIL 179
|||||
902 GCGCTCGATTCATTAATTCGACCGAGATGCGAGATGCGTGGAAAATT 951
|||||
179 EGLNHSGLUVALGLYILETHRTHRVALTYRVALTHRHSASPLINGLUG 196
```

```

1952 ACAGGCGAAACATCAGCGCACCATGTCCTTATTCCACAGATCTGATG 1001
196 IuAlaMetAlaIleSerAspGlnIleAlaValMetLysAspGlyValIle 212
1002 AACCCATGCTAATGGGACCGAATTCATATGCAAAATGGTGAAGTG 1051
213 GlnGlnIleGlyArgProLysGlnLeuTyrHisLysProLysAsnGluPh 229
1052 GTACAGGTGGCACACCGGATGAAATTCATATATCCGCGCAATGATTA 1101
229 eValAlaIlePheIleGlyArgThrAsnIle 239
1102 TGTCGTAACCTCTCCGTCGCTGATATT 1132

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-402-804-5

seq_documentation_block:

```

Sequence 5, Application US/08402804
Patent No. 5874300
GENERAL INFORMATION:
APPLICANT: Biasez, Martin J.
APPLICANT: Pei, Zhiheng
TITLE OF INVENTION: Campylobacter Jejuni Antigens, And
TITLE OF INVENTION: Methods For Their Production And Use
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: OSTROLENK, FABER, GERB & SOFFEN
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,804
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/114,420
FILING DATE: 30-AUG-1993
APPLICATION NUMBER: US 08/112,387
FILING DATE: 27-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,928
FILING DATE: 08-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/612,330
FILING DATE: 13-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gray III, William O.
REGISTRATION NUMBER: 30,944
REFERENCE/DOCKET NUMBER: P/1261-13
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 881 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
FEATURE:
NAME/KEY: CDS
LOCATION: 132..857
US-08-402-804-5

```

alignment_scores:
 Quality: 349.00 Length: 237
 Ratio: 2.115 Gaps: 2
 Percent Similarity: 69.620 Percent Identity: 31.646

alignment_block:

US-09-769-787-162 x US-08-402-804-5 ..

Align seg 1/1 to: US-08-402-804-5 from: 1 to: 881

```

4 ILeLysIleIleAsnAlaLysLysIleTyrHisAspValProValIleG1 20
135 ATGAATTAATAAAATGTAAACAATACTAGCAAGCATCATCATGTCTAAA 184
20 uAsnLeuAsnIleThrIleProLysGlnSerLeuPheThrLeuGlyA 37
185 GATATTTAATCTTGTGTTAAAGAGGTGAAGCTGTTATTATATAGTTC 234
37 IAserGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsn 53
235 CAAGTGAAGAGTGAAGAAAGTACACATATCCGTTGCATGAAATGGCTTGA 284
54 SerIleGlnGlyGlnPheTyrPheAspAspThrLysIleAsn.. 69
285 GAAGTTAGTTACAGAGAGGTCGTAGTTAACAAATCTGTTTAAATCATTA 334
70 .....MetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84
335 AAATTAATAATGAATTTGCCGAAATATTTGTGCAATGGTTTTCACCAT 384
84 YrAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
385 TTAATTTATATCCACATATGACGCTTTTGCAAAATTTGACCTTAGCTCCA 434
101 MetGln...LysLysValProLysGlnGlnLeuIleGlnGlnThrAsn 116
435 ATGAACCTTCAAAAAAATCTAAAAAAGAGCTGAAGAAACACCTTTTAA 484
116 sTyrLeuGlnLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAsp 133
485 GTATTTAAAGGTTGTAGGTTGCTGATTAAGCAATGTTTATCCAGCAA 534
133 YsLeuSerGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAla 149
535 CCCTTTCAGGTGACACACACACACGCTTCTATAGCAATCATCTTTGT 584
150 ValAsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGlu 166
585 ACTAAAAAACCTATATTTTATTTGATGAACCTTACGCCCTTGATCC 634
166 aLysLeuArgLeuAspMetArgGlnAlaIleArgGlnIleGlnHisGlu 183
635 AGAAACCATCAAGAGGTTTATGATGTAAGAAAGAAATTTCAATCATA 664
183 aLysIleThrValTyrValThrHisAspGlnGlnGlnAlaMetAla 199
685 GCATATCTACCATGTCGTTGTTACACACGAATGGGTTTGCAAAAGAA 734
200 ILeSerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIle 216
735 GTAGCAGATAGATTTATTTATTTATGAGAGATGCTATTTGTGGAAGAA 784
216 YArgProLysGlnLeuTyrHisLysProAlaAsnGluPheValAlaThr 233
785 TATTCCTAGTGAATTTTCTCAATCCAAAAGCTGAAGAGCGGACTCT 834
233 heIleGlyArg 236
835 TTTTAGGAAA 845
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-08-919-573-3
seq_documentation_block:

```

```
; Sequence 3, Application US/08919573
; Patent No. 6346392
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; TITLE OF INVENTION: ATP-BINDING PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,573
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50597
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-919-573-3
;
; alignment_scores:
; Quality: 345.00 Length: 238
; Ratio: 2.091 Gaps: 3
; Percent Similarity: 69.328 Percent Identity: 31.513
;
; alignment_block:
; US-09-769-787-162 x US-08-919-573-3 ..
;
; Align seg 1/1 to: US-08-919-573-3 from: 1 to: 732
;
; 4 Tlelystlleasnalalyslysleryrlhsapvalprovalillegl 20
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 16 ATTAATAATTGAATAATTACATTAATCTTTGAGGAGATGAATGATGAA 65
; 20 uasnleuasnllethrilleprolysglyserleuphetherleuengly 37
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 66 GGGCATCACTCGAGATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 115
; 37 laserglycysglylysthrthrleuleuarqmetilleaglypheasn 53
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 116 CTTGAGGAGAGTGGAAATCTACCTTGCTCGCTCATGAATTTGTTGAG 165
; 54 serlleuglyglyluhphetyrphesapsphirlysle..... 67
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 166 GAAGCAACCAAGGAGGAGGATATCTTTGAGGAGCGATATTACGACAA 215
; 68 .....Asnasmkeltuproserlysarqasnilleglymetvalpheglna 83
; |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 216 GAAGAATGACTCTTTTGGCATGCGTGAAGAGATGGGAGTGTTCACAC 265
```

```
83 snTyralallepheprohisleuthrvalargaspasvalalaphgly 99
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 266 AATTAACCTCTTCTTAATATGATGATGATGATGATGATGATGATGATG 315
; 100 leuuetglnlyslsvalprolysglygluileuilegln...glnthrns 115
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 316 CCTATCAAGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 365
; 115 nlystlyrleuileuemetglnilleaglnlytvalasparglysppca 132
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 366 AAGCTTTTGGAAAAAGTTGGTTTGGCCAGTAAGCAGACGCTTATCCAC 415
; 132 splysleuserglyglynglnnglnargvalthrleualacysalaleu 148
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 416 AGAGCTCTCGGTGGGAGCAACAGGATTCATCGCCGCGGTGG 465
; 149 Alavalasnploservalleuleumetaspluproleuserasnlengl 165
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 466 GCTATGGAACCAAGATGTTTGTCTTTGACGACCAACCTTCAGCCTTGA 515
; 165 uAlalysleuarglyleuaspmetarqlnAlaileargluileglnhsg 182
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 516 TCCTGAATGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 562
; 182 luvalglylethrthrvalthrvalthrhisaspgingluAlaIleMet 198
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 563 AGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 612
; 199 AlalieserapnglnleAlaIleMetlysaspllyallieglnnglnl 215
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 613 GAGTGGCAGATCGTGTATCTTATGACAGACCGGTGTGTGTGAGA 662
; 215 eglYArgProlysglyleuThrHislyslsProAlaasngluPhValaIar 232
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 663 TGGGACACCTGAGAGATTTTGAACAACCAACCAACCAACCAACCAAC 712
; 232 hrPheIleGlyArg 236
; :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
; 713 ATTTCTTGAGTAA 726
;
; seq_name: /cgn2_6/prodata/1/lna/6B_COMB.seq:US-08-919-573-1
;
; seq_documentation_block:
; ; Sequence 1, Application US/08919573
; ; Patent No. 6346392
; ; GENERAL INFORMATION:
; ; APPLICANT: Burnham, Martin, Karl Russel
; ; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; ; TITLE OF INVENTION: ATP-BINDING PROTEIN
; ; NUMBER OF SEQUENCES: 6
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Dechert Price & Rhoads
; ; STREET: 997 Lenox Drive, Building 3, Suite 210
; ; CITY: Lawrenceville
; ; STATE: NJ
; ; COUNTRY: USA
; ; ZIP: 08543
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Diskette
; ; OPERATING SYSTEM: DOS
; ; SOFTWARE: FASTSEQ for Windows Version 2.0
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/919,573
; ; FILING DATE:
; ; CLASSIFICATION: 530
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER:
; ; FILING DATE:
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Bloom, Allen
; ; REGISTRATION NUMBER: 29,135
; ; REFERENCE/DOCKET NUMBER: P50597
```

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609-520-3214
: TELEFAX: 609-520-3259
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 735 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
US-08-919-573-1

alignment_scores:
      Quality: 345.00      Length: 238
      Ratio: 2.091      Gaps: 3
      Percent Similarity: 69.328      Percent Identity: 31.513

alignment_block:
US-09-769-787-162 x US-08-919-573-1 ..

Align seg 1/1 to: US-08-919-573-1 from: 1 to: 735

4 11elysileleasna1alyslyletryhisaspvalproval1leg1 20
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
16 ATAAATTTGAAATTTACATTAATCTTTGCAAGAGATGATTTGAA 65
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
20 uasneusnilethrlleprolysglyserleuphetrlleu1lya 37
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
66 GGGCATCAACCTCGAGATTAAAGAGAGAACGTTGCTTATCATCGGTC 115
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
37 laSerGlycylslysthrthrlleu1argmetlealaglypheasn 53
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
116 CTTGAGGAGTGGAAATCTACCTGCTTCTGCTATGCAATTGTTGAG 165
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
54 Serilegylglyluetheryrpheasprthrlyslle..... 67
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
166 GAAGCAACCAAGGAGGTATCTTTGAGGAGTCGATATTACGACAA 215
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
68 ...Asnasmetgluproserysargasnileglymetvalpneqla 83
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
216 GAAGATGACCTGTTGCCATCGTGAGAAAGTGGCATGTTTCCAA 265
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
83 snTyrallellepheprohisleuthrvalargaspasvalalaphegly 99
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
266 AATTTAACCTCTTCCATATATGACTGATTTGAAATATTTACCTGTA 315
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
100 leumetglylsva1prolysglu1leu1leg1n...Glnthras 115
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
316 CCTATCAAGACCAAGGTGAAGCAGGGAAGTTGCTGACAAAGAGCCCA 365
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
115 nlystlyrleuglu1leumetgl1lealag1nlyr1alasparglysp1ca 132
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
366 AAAGCTTTTGGAAAAAGTTGGTTTCCAGATTAAGCAGACGCTTATCCAC 415
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
132 splyleu1sergly1ng1ng1nargval1thrlleu1lacy1alaleu 148
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
416 AGAGTCTGTGGGTGGGACAGCAACAGGATTCGCGGTGGGTGG 465
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
149 AlaValasProserValleu1leumetasp1upro1leu1SerAsnleu1 165
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
466 GGTATGGAACCAAGATGTTTGTCTTTGACGAAACCACTTACGCTTGA 515
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
165 uAlalysleuargleu1aspmetarg1nAlalearg1u1leg1n1h1sg 182
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
516 TCCTGAATGTAGGTGAGGTATGCTGTATGCAAGACCTTGCC...A 562
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
182 luValGlylethrlthrval1tyrval1thrlhisasp1ng1lu1u1amet 198
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
563 ACTCAGGGAGTACTATGTTATCGTAACCAAGATGAGATTGGACCTTGC 612
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199 Ala1leSerasp1n1leAlaVal1metly1asp1g1yVal1leG1n1l 215
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
613 GAGTGGCAGATCGTGTATCTTTATGCGAGACGGTGTGTTGTAAGA 662
215 eGlyargProlysglu1leu1tyrhislysp1ro1alasn1upheval1at 232
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
663 TGGACACCTGAGCAGATTTTGAACAAACCAAGAACAGCGACTTAAG 712
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
232 hrPhe1leGlyarg 236
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
713 ATTCTCTGAGCTAAG 726

seq_name: /cgn2_6/plodata/1/ina/6B_COMB.seq:US-08-858-207A-49

seq_documentation_block:
: Sequence 49, Application US/08858207A
: Patent No. 6348328
: GENERAL INFORMATION:
: APPLICANT: Black, Michael
: APPLICANT: Hodgson, John
: APPLICANT: Knowles, David
: APPLICANT: Nicholas, Richard
: APPLICANT: Stodola, Robert
: TITLE OF INVENTION: No. 6348328bel Compounds
: NUMBER OF SEQUENCES: 552
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmltnKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/858,207A
: FILING DATE: 09-MAY-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/017670
: FILING DATE: 14-MAY-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Gimm1, Edward R 891
: REGISTRATION NUMBER: 38,891
: REFERENCE/DOCKET NUMBER: P50475
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-4478
: TELEFAX: 610-270-5090
:
: TELEX:
: INFORMATION FOR SEQ ID NO: 49:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1460 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-08-858-207A-49

alignment_scores:
      Quality: 339.50      Length: 202
      Ratio: 2.219      Gaps: 3
      Percent Similarity: 75.743      Percent Identity: 34.653

alignment_block:
US-09-769-787-162 x US-08-858-207A-49 ..

Align seg 1/1 to: US-08-858-207A-49 from: 1 to: 1460

2 SerGlu1le1ys1le1leasna1alysly1letryhisaspvalprova 18
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
367 TCTATGTGAATTAAGAAATATCAATAAAGTCTTGGAGACAAACAAT 416
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

18 *lileg*uanleuAsn¹leth¹l¹leProlysgly¹serleuPheh¹leu¹ 35
417 CCGTGTGATTTTCAGCTCTAAGATTCTCTGAAAGCAAACTCGGTATCG 466
35 *eug*l¹Alasergly¹cysgly¹sth¹rh¹leuAeAr¹gmet¹leA¹gly 51
467 TTGGACCTTCCTGGTAGAGTAGAACAACTCTTTACGATCTTCGCGGT 516
52 *Pha*ns¹er¹leGln¹gly¹ylu¹b¹u¹he¹yr¹Pha¹s¹ps¹ph¹lys¹Ile¹as 68
517 CTTGAAACCTTGATTCACGGCAAACTTTTATATGACACACTTTAGA 566
68 *nas*met¹glu¹Pro¹ser¹Lys¹Arg¹Asn¹...*Ile*gly¹met¹Val¹Ph¹e¹ln¹Asn¹ 84
567 GCTGGATGAATTGCAGAAACCCAACTCTACTGSGATTTGTCTCCAAAGAT 616
84 *yl*Al¹lePhe¹Pro¹His¹leu¹rh¹Val¹Arg¹Asn¹Val¹Ala¹Phe¹gly¹leu 100
617 TTCACTATTTCCTCATCTATCATAGTTGGAAAAATTTCACCTTTATCGCT 666
101 *Met*glu¹Lys¹Lys¹...*Val*¹Pro¹Lys¹glu¹leu¹leu¹Ile¹gln¹h¹rh¹Asn¹ly 116
667 GTGAGACCATGGGAATGATGAGCAGAGAAAGGCTGAGAAAGGCGAGTGG 716
116 *s*Tr¹leu¹glu¹leu¹Met¹glu¹leu¹Ala¹gln¹Tr¹Ala¹s¹par¹gly¹Pro¹Asp¹ 133
717 ACTCTTGGAAACAGTTAAGACTGTGGAGGACACGAGAGCTCTATCCCTTTC 766
133 *ys*leu¹ser¹gly¹gln¹gln¹gln¹h¹rh¹Val¹rh¹leu¹Ala¹cy¹Ala¹leu¹ 149
767 CACTATCTGTGGGCCAAAGCAGCGGGGTGCTTTGGCGCGGTCTATGATG 816
150 *Val*Asn¹Pro¹ser¹Val¹leu¹leu¹Met¹Asp¹lu¹Pro¹leu¹Ser¹Asn¹leu¹glu¹ 166
817 ATTGACCCTCAGAAATCATTTGGCTACAGTAAACCAACTTCCCTCCGATCC 866
166 *Al*ys¹leu¹Arg¹leu¹Asp¹Met¹Arg¹Ala¹le¹Arg¹lu¹Ile¹gln¹h¹se¹ly 183
867 AGAATTAACGTTGGAGAGTGAGCAACCTATATCTTGCAA...*AA*TGGGAAC 913
183 *Al*g¹ly¹le¹rh¹rh¹Val¹Val¹Tr¹h¹rh¹His¹Asp¹gln¹glu¹Ala¹Met¹Ala 199
914 TTGGGATGACCAAGATGTGGTTACCATCCATGATTTGGCAAGTTTGGCTGAAT 963
200 *I*le¹ser 201
664 ATCGCC 969


```

509 GTTCCAGCCAAAGCAGCAATATACCATGTGTTCCAGAAATTACGCCTT 460
86 ephProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnL 103
   :::::::::::::::::::: ::::::::::::::::::::
459 GTATCCGATATGAGCGTTTACGACAACATGCGCTTGTCTGAAGATGC 410
103 ySLysValProLysGlnGluLeuIleGlnGlnThrAsnLysTyrLeuGlu 119
   :::::::::::::::::::: ::::::::::::::::::::
409 AAAAATCTGCCAAAGAGCTGATGTGATGAGCGGTGCAACTGGCGCGCA 360
120 LeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerG 136
   :::::::::::::::::::: ::::::::::::::::::::
359 ATTCTGCGCTGCGTGAATGACCTGAAACGTAAAGCCGGGCGCTTCCGG 310
136 yGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnPro 153
   :::::::::::::::::::: ::::::::::::::::::::
309 CGGGCAACGTCAGCGAGTGGCGCTTGGCGGCGCATCTACCGCAAGCGG 260
153 eValLeuLeuMetAspGluProLeuSerAsnLeuGlnAlaLysLeuArg 169
   :::::::::::::::::::: ::::::::::::::::::::
259 GCGTGTATTATGATGATGAACCGCTCTTACCTGTGATGCCAAGCTGCC 210
170 LeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleTh 186
   :::::::::::::::::::: ::::::::::::::::::::
209 GTGCAATGCGCGCAGAGATCAGCAGCTGCATCGAAGAACTGAAACACCA 160
186 rThValTyrValThrHisAspGlnGluGlnAlaMetAlaIleSerAsp 203
   :::::::::::::::::::: ::::::::::::::::::::
159 CATGATTAAGTACGACCCAGATCAGACCGAAGCGATGACATGGGAGCG 110
203 InIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 219
   :::::::::::::::::::: ::::::::::::::::::::
109 GGAATGTGATATATGAAAGACGCGATGTTCAGCAAGTGGTGGCGCGCAA 60
220 GluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGly 235
   :::::::::::::::::::: ::::::::::::::::::::
59 ACCGTTTATTAACCAACCGCGAATATGTTGTTCCGATTTATTTGA 12

```

seq_name: gb_est2:BG855631

seq_documentation_block:

LOCUS BG855631 720 bp mRNA linear EST 29-MAY-2001

DEFINITION 1024043612.x2 C. reinhardtii CC-1690, normalized, lambda zap II

ACCESSION Chlamydomonas reinhardtii cDNA, mRNA sequence.

VERSION BG855631

KEYWORDS BG855631.1 GI:14236815

SOURCE EST.

ORGANISM Chlamydomonas reinhardtii.

REFERENCE Chlamydomonas reinhardtii

AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,

McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.,

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; Project phase 2

JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu

FEATURES

source

1. /720

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, normalized, lambda zap

II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:

XhoI. This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO₂ and HS medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 157 a 215 c 180 g 166 t 2 others

alignment_scores:

Quality: 468.50 Length: 235

Ratio: 2.617 Gaps: 3

Percent Similarity: 76.170 Percent Identity: 41.702

alignment_block:

US-09-769-787-162 x BG855631/rev ..

Align seg 1/1 to reverse of: BG855631 from: 1 to: 720

63 ASPSPRTHRLYSILEASNASIMETGLUPROSERLYSARGASNIIEGLYME 79

710 GATGGCGTCGATNTGTACAGATTCCGCCCTTACCTGCGCCCATCATATAT 661

79 tValPheGlnAsnTyrAlaIlePheProHisLeuThrValArgAspAsn 96

660 GATGTTTCAGTCTTACGCGCTGTTCCCATATATGACCTGGAGACAGAA 611

96 aAlaIlePheGlyLeuMetGlnLysLysValProLysGlnGluLeuIleGln 112

610 TCGCTTTGGCTCGAAGACAGCAAACTACCGAAGCCGGAATTTGCCAGC 561

113 GlnThrAsnLysTyrLeuGlnLeuMetGlnIleAlaGlnTyrAlaAsp 129

560 CGGGTCATGATGATGCTCGCGCTGTCATATGACAGAGTTCCGCAACG 511

129 gLysProAspLysLeuSerGlyGlyGlnGlnGlnArgValThrLeuAla 146

510 CAAACCGCATCAGCTTCCGCTGTCACAGCAACGTCGTCGCCCTGGGCC 461

146 yAlaIleAlaValAlaAspProSerValLeuLeuMetAspGluProLeuSer 162

460 GAACGCTTGGCAGACGCCGGAACCTATTACTGCTCGATGACCGCATGGGC 411

163 AsnLeuGlnAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGlu 179

410 GCGCTGATTAAGAAAGCTCCGTCGACAGATGACGCTTGAAGTGGGATAT 361

179 eGlnHisGluValGlyIleThrThrValTyrValThrHisAspGlnGlu 196

360 TCTGAGCGCGCTCGCTGATGATGTGTGATGATGATGATGATGATGAT 311

196 lValIleMetAlaIleSerAspGlnIleAlaValMetLysAspGlyVal 212

310 AGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261

213 GlnGlnIleGlyArgProLysGlnLeuTyrHisLysProAlaAsnGlu 229

260 GTCAGATGTCGCAACCGAGAGATGATGACAGATGATGATGATGATGAT 211

229 eValAlaThrPheIleGlyArgThrAsnIleIleProAlaAsnLeuGlu 245

210 TAGCGCGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 161

246 LysArgSerAspGlyAlaTyrIleValPheSerAspGly.....TyrAl 260

160 GAGCGTCAGGAAGATGGCTGCTGATTCGCGCGGCGCTGTCATATTC 111


```

MEDLINE      21376150
COMMENT      Contact: Wing RA
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: rwing@clemson.edu
              Class: BAC ends
              High quality sequence stop: 809.
              Location/Qualifiers
FEATURES
  source      1. 874
              /organism="Bradyrhizobium japonicum"
              /strain="USD110"
              /db_xref="taxon:375"
              /clone_lib="B. japonicum BAC library"
              /lab_host="E. coli"
              /note="Vector: pindigo536. Site_1: HindIII"
BASE COUNT   146 a      268 c      281 g      175 t      4 others
ORIGIN
alignment_scores:
  Quality:    417.00      Length:    272
  Ratio:      2.206      Gaps:      6
  Percent Similarity: 69.485      Percent Identity: 36.765
alignment_block:
  US-09-769-787-162 x AZ934163/rev ..
Align seg 1/1 to reverse of: AZ934163 from: 1 to: 874
41  G1yLysThrThrLeuLeuArgMetIleAlaG1yPheAsnSerIleGluG1 57
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 GGGCAAAACCCACGCGATGATAGTGGCTGGTTGAGTCCGACCATG 811
57 yG1yLyuPheTyTPhAspThrLysIleAsnMetGluProSerL 74
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
810 GGAAGGATTTTCGCAAGACGCCGATCTAAACATCCCGCCGACAA 761
74 ySArGAsnIleGlyMetValPheGlnSerTyRAlaIlePheProHISLeu 90
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
760 AGCGGCGACATCGCATGTGTTCAGATTATGCTTTGTGTCGCGACATG 711
91 ThrValArGAspAsnValAlaPhe.G1yLeuMetGlnLysValProL 107
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
710 CGATGCCGGAATAATATNCTTCGCCGCTC...GTCGCGAATATCACA 664
107 ySgLUgluLeuIleGlnGlnThrAsnLysTyRLeuGluLeuMetGlnIle 123
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 AGCCGNAAGCGCAGAACGCCGTCAGNCGCGCTCCCATGATCAGATGATG 614
124 AlAG1yTyrAlaAspArgLysProAspLysLeuSerGlyGlnGlnG1 140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 AAACCCCTGGCCACCGCCGCGCGGCGAGCTGTCGCGGTGACGACGA 564
140 nArGValThrLeuAlaCysAlaLeuAlaValAsnProSerValLeuLeu 157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
563 GCGCGTGGCCTCGCGCCGCGGCTGTTTCATCCGACATCGCATGCTGTA 514
157 eLAspGluProLeuSerAsnLeuGluAlaLysLeuArGLeuAspMetArG 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
513 TGGAGAGCGCCCTTGGCGCTGACAAAGCGCTGCGGAGGAGATGCCAA 464
174 G1AlaAlaArgGluIleGlnHISGluValG1yIleThrThyValTyRva 190
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 CTGGAATCAAGCAACTGCACGAGACGATGGGCGCTCGCTACGTACGT 414
190 lThyHISAspGlnGluGluAlaMetAlaIleSerAspGlnIleAlaValM 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 CACCCAGATCAGATGAAGCGCTCACCATGTCGAGACCGGATGGCGCTGT 364
207 eLysAspGlyValIleGlnGlnIleGlyArGProLysGluLeuTyRHIS 223

```

```

363 TCAACGACGCGATCGTCACAGATGCACAGGCCGCGCTGTATGAG 314
224 LysProAlaAsnGluPheValAlaThrPheIleGlyArGThrsnIleI 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 CATCGGTGAACAGCTTCGTGCGCTCACTTATCGGCGAGAACATGTGCT 264
240 eProAlaAsnLeuGluLys...ArgSerAspGlyAlaTyRILEValPheS 256
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 GCGCGGACCGCTGACAGCGTGCAGCAAGACTATTGCCGCGCGCTGG 214
256 eArSpGlyTyRAlaLeuArgMetProAlaLeuAspGlnValG1yGluGln 272
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 CGGCGCGCGCGCGCTACCGCACGAGCGGGTCAATGATCAGCGCGCGC 164
273 AlA...IleHISValSerIleArGProGluGluPhe...IleLysAspG1 287
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 GCATCGACCTCCCTGTGTCGCGCGCGGAGCGCTGCTGTGCCGGA 114
287 uSerGlyAspIleGluGly.....ThrIleArgAsps 298
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 CGGACACTCCAGCGAAGACCGAACCAGCGCTGCCGCGAGGTGCAGACA 64
298 eRValTyRLeuGly 302
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 CCATCTATCTCTCGGC 50
seq_name: gb_est1:AV594537
seq_documentation_block:
LOCUS      AV594537      454 bp      mRNA      linear      EST 27-NOV-2001
DEFINITION AV594537 Bos taurus cartilage fetus Bos taurus cDNA clone
E1CA014C12 3', mRNA sequence.
ACCESSION AV594537
VERSION    AV594537.1 GI:9710934
KEYWORDS   EST.
SOURCE     cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 454)
            Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
            Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
            Nucleic Acids Res. 29 (22), E108 (2001)
21570554
JOURNAL    MEDLINE
COMMENT    Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shriakawa Institute of Animal Genetics
            Odokura, Nishigo, Nishi-shriakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES
  source      1. 454
              /organism="Bos taurus"
              /db_xref="taxon:9913"
              /clone="E1CA014C12"
              /clone_lib="Bos taurus cartilage fetus"
              /tissue_type="cartilage"
              /dev_stage="fetus"
              /lab_host="DH10B"
              /note="Vector: pZL1, Site_1: SalI; Site_2: NotI; Poly A
              was deleted from a NotI site"
BASE COUNT   98 a      131 c      124 g      101 t
ORIGIN

```

alignment_scores:

Quality: 404.00 Length: 151
Ratio: 3.311 Gaps: 0
Percent Similarity: 80.795 Percent Identity: 50.331

alignment_block:

US-09-769-787-162 x AV594537/rev ..

Align seg 1/1 to reverse of: AV594537 from: 1 to: 454

```

77  TLEGLYMETVALPHEGLNASNTYRAIALIIEPHEPROHISLEUTHVALAR 93
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453  GTTGGATGGTGTTCAGTCTTACGGCTCTATCCACCCTGCAGTAGC 404
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93  GASPARNAVAlAlaPheGLYleuMetGLInLysValProLYSGluGLInL 110
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403  AGAAACATGTCATTGGCTGAACCTGGCTGGCGCAAAAAGAGGTGA 354
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110  euLleGLInLThrAsnLysTYrLeuGLInLLeuMetGLInLleAlaGLInTYr 126
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353  TTAAACACGCCGTTAACAGAGTGGCGAAGTGTACACACTGGCGCATTTG 304
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127  AlAspArgLysProAspLysLeuSerGLYGLInGLInGLInArgValTh 143
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303  CTGGATCGCAACCGAAAGCGCTCTCGGTGTCAGCGTCAGCGTGGCGC 254
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143  rLeuAlaCysAlaLeuAlaValAsnProSerValLeuLeuMetAspGLInP 160
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253  GATTGGCCGTCAGCTGTGGCGGACGCAACCGTATTTTGTCTGATCAAC 204
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160  rLeuSerAsnLeuGLInLAlaLysLeuArgLeuAspMetArgGLInAlaIle 176
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203  CGCTCCCAACCTCGATGCTGCACCTGGTGTGCAATGCTATCGAATC 154
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177  ArgGLInLleGLInLsGLInValGLYIleThrThValTYrValThrHisAs 193
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153  TCCCGCTGCTAATACGCCCTGGCGCCGACAAATGATTACTCTACCCACGA 104
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193  pGLInGLInLAlaMetAlaIleSerAspGLInLleAlaValMetLysAsp 210
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103  TCAGGTCGAACGATGACGCTGGCCGCAAAATGCTGGTCTGGACCGC 54
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210  lValLleGLInLleGLInLArgProLYSGluLeuTYrHisLysProAla 226
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53  GTGGCGGCGCGCAGGTTGGGAAACGCTGAGCTGATCCACTATCCGCA 4
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227  Asn 227
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3  GAC 1
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_gss:AZ935459

seq_documentation_block:

LOCUS AZ935459 545 bp DNA linear GSS 24-APR-2001
DEFINITION BJ-Ba000H18r B. japonicum BAC library Bradyrhizobium japonicum
genomic, DNA sequence.
ACCESSION AZ935459
VERSION A2935459.1 GI:13778216
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum.
ORGANISM Bradyrhizobium japonicum.
Bacteria: Proteobacteria; alpha subdivision, Rhizobiaceae group;
Bradyrhizobium group; Bradyrhizobium.
1 (bases 1 to 545)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Judd,A., Goloechea
J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 502.
Location/Qualifiers

FEATURES

source
1..545
/organism="Bradyrhizobium japonicum"
/strain="USD110"
/db_xref="taxon:375"
/clone_lib="B. japonicum BAC library"
/lab_host="E. coli"
/note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT 86 a 167 c 197 g 93 t 2 others
ORIGIN

alignment_scores:

Quality: 385.00 Length: 164
Ratio: 3.080 Gaps: 0
Percent Similarity: 76.220 Percent Identity: 46.341

alignment_block:

US-09-769-787-162 x AZ935459/rev ..

Align seg 1/1 to reverse of: AZ935459 from: 1 to: 545

```

67  TlAsnAsnMetGLInProSerLysArgAsnLleGLYMetValPheGLInAs 83
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492  GTCACTTACATGCCCGCCGACAGCGCGACATCGCGTCTATTCCAGAA 443
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83  nTYrAlaIlePheProHISleuthrValArgAspAsnValAlaPheGLYl 100
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442  CTACGGCGCTGTCCCGCATATAGACGCTGGCGGAACGTCGCTTCCCN 393
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100  euMetGLInLysValProLYSGluGLInLleuLleGLInLThrAsnLys 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392  TGGCGCGCGCCCACTGCCGAAAGCGAGCGCGGCAAAAGTGGCGCG 343
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117  TYrLeuGLInLLeuMetGLInLleAlaGLInTYrAlaAspArgLysProAsp 133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342  GCCCTCGCATGTGCTGGCTTGGCGCTACGAGGAACGCGCATCCGCA 293
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133  sLeuSerGLYGLInGLInGLInArgValThrLeuAlaCysAlaLeuAlaV 150
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292  GCTCTCGCGCGCAGCGCCAGCGCGGTGGCGCACGGCCCATGATCT 243
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150  AlAsnProSerValLeuLeuMetAspGLInProLeuSerAsnLeuGLAla 166
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242  TCAGAGCGCGCTGATCTCATGAGACGACCGCTATCCGCTCGACAG 193
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167  LysLeuArgLeuAspMetArgGLInAlaIleArgGLInLleGLInHisGLu 183
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192  CAGCTGGCGCAATCATGACATGACGCTCGCGCGCTGACCGGCGCAT 143
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183  lGLYIleThrThValTYrValThrHisAspGLInGLInLAlaMetAlaI 200
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142  CGGCGCACCATCATCTACGCCATGATGATGATGATGATGATGATGATGAT 93
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200  lSerAspGLInLleAlaValMetLysAspGLYAlaLleGLInGLInLleGLY 216
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92  TGAAGCGACCGCTGCTGATCATGAGACGCGGATGATCATCATGATGATG 43
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217  ArgProLYSGluLeuTYrHisLysProAlaAsnGLInPheVal 230
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42  GAGCGGCGACGTCGACGACATCCCGCGGATCTCTTCGTC 1
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_gss:AO012091

seq_documentation_block:

LOCUS AO012091 670 bp DNA linear GSS 04-JUN-1998
DEFINITION 531pG077070297 Cosmid library of chromosome II Rhodobacter

```

ACCESSION   spheraoides genomic clone 531PG0770297, DNA sequence.
VERSION     A0012091
KEYWORDS    A0012091.1 GI:3177046
SOURCE      GSS.
ORGANISM    Rhodobacter sphaeroides.
            Rhodobacter sphaeroides
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodobacter.
REFERENCE   1 (bases 1 to 670)
            Choudhary, M., Mackenzie, C., Mouncey, N., Weinstock, G. M. and Kaplan,
            S.
            ResDB, the Rhodobacter sphaeroides Genome Database
            Unpublished (1998)
            Contact: Choudhary, M.
            Department of Microbiology and Molecular Genetics
            University of Texas Medical School
            6431 Fannin Street, Houston, TX 77030, USA
            Tel: 713 500 5437
            Fax: 713 500 5499
            Email: madhuetumg.med.utx.tmc.edu
            Seq primer: PBLuescript T7
            Class: shotgun.
FEATURES
    source          Location/Qualifiers
                     1..670
                     /organism="Rhodobacter sphaeroides"
                     /strain="2.4.1T"
                     /db_xref="taxon:1063"
                     /clone="531PG0770297"
                     /clone_1lb="Cosmid library of chromosome II"
                     /lab_host="E. coli S17-1"
                     /note="Vector: pLA2917"
BASE COUNT      120 a      232 c      203 g      115 t
ORIGIN
alignment_scores:
    Quality:      377.00      Length:      225
    Ratio:         2.299      Gaps:         4
    Percent Similarity: 72.889      Percent Identity: 38.222
alignment block:
US-09-769-787-162 x A0012091 ..
Align seg 1/1 to: A0012091 from: 1 to: 670
5 LyslleileasnlalyslyllyrHlaspvalProvalllegluas 21
|||||::: ||| ||| ::::|::|::|
7 AAGATCGTG.....AAGCGGATATGGCGGATGAGAGCCATCCACGG 47
21 nleuanslletHrllleProlysglySerleuphetHrleuenglalas 38
:::|::|::: ::| |::|::|::|::|
48 CCGTCATCTCGAGTGGAGAGCGGCGAGCTTCGTGGCGCCCT 97
38 eGlyCYsglylYstrHrthlrleuauAGmetlleAlaglyPheasnsr 54
|||||:::|::|::|::|::|::|::|::|::|::|::|::|::|
98 CGGGCTCGCGGAAATCCACATCGTCGCGGATGATTTGCCGGCTGAGGAC 147
55 lleglglglygluPhetYrPheaspasPthrlYlIsleasAsmetG1 71
|||::|::|::|::|::|::|::|::|::|::|::|::|::|
148 ATCAGCGGCGCCATATGCGATCGCGGCGGCGGTGGAACGATATGCA 197
71 uProserlysaArgasnllleglymetValPheglinasnYrAlalePhep 88
|||||:::|::|::|::|::|::|::|::|::|::|::|::|::|
198 GCCCAAGGCGCGCGACGTGGCGATGTCTTCAGAGATTACCGCGCTATC 247
88 rHslleuThrValArgaspasnsValAlaPheglyLemegLlnslyls 104
|||||:::|::|::|::|::|::|::|::|::|::|::|::|::|
248 CGCAATAGACGCTGCGCGACAACATCGGCTTCGGGCTAAATCGCGCGC 297
105 ValProlysgluGlulenuileglnGlnThrasnlyTYrleuGluleu 121
||| ||| ::::|::|::|::|::|::|::|::|::|::|::|::|::|
298 GAGCCCGCGAGCATTCGGAAGAACTCAGAGAAAGCCGCGCGCATCT 347

```

```

121 tctnlttcaatgtttatgaaacgcggccgcccgcagcaccgaagtccgaggcgccg
348 ccagcttcgacgacctctgcacgccggccggccgacctctcgaggcgccg
138 lnglncnglnrvgaltnrlleualacysalaaleualaavalasnpsoervval 154
||||| ||| ::|||::|
398 agggctcagcccgcttgcatctggccccccctccctccgacccaaccaagtc 447
||||| ||| ::|||::|
155 leulenmetaspcluproleuserrasnleulgualalauleargleuas 171
::::: |||::| | |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
448 ttccctcttgacaaaaccctctccaacctccaacccacgctgagggtccca 497
||||| ||| ::|||::|
171 pmetarginalallearglullegnlnhsuglvalglyllethr-thrv 188
||||| ||| ::|||::|
498 aaagccgac. cagacttatccgggtcccccgatttgccgcccgacgacga 546
altrvrtalthrhisaspnglnglu. alamelalaleserapslnl 204
::::::::: |||||::|::| |||||::|::|::|::|::|::|::|::|::|::|::|
547 tctatrttaaccgccgatcaggttcaggcgatcaccttgccgacggcgr 596
204 ealaval... Metlysaspcllyallleglnlleglaiaiprolsg 220
::::: ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
597 tggttgttgtccgccgacgacatcaccgacgacgccctgcctcc 646
||||| ||| ::|||::|
220 luleu. tyrhislvsprola 226
::||| ||| ::|||::|
647 aactcttacaacagcgccctcc 667
seq_name: gb_gss:CNS07HFA
seq_documentation_block:
LOCUS CNS07HFA 557 bp DNA linear GSS 02-OCT-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 26P08 of library NotreDame1
from strain PE8T of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL610984
VERSION AL610984.1 GI:15917169
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
REFERENCE 1 (bases 1 to 557)
Genoscope. Direct Submision Submitted (01-OCT-2001) Genoscope - Centre National de Sequenace : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
2 (bases 1 to 557)
Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H. Direct Submision Submitted (01-OCT-2001) BMVI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES
source location/Qualifiers
1..557
/organism="Anopheles gambiae"
/strain="PE8T"
/db_xref="taxon:7165"
/clone="26P08"
/clone_lib="NotreDame1"
/note="end : SP6"
BASE COUNT 106 a 147 c 175 g 129 t
ORIGIN
alignment_scores: Quality: 371.50 Length: 161
Ratio: 2.972 Gaps: 2

```

Percent Similarity: 77.640 Percent Identity: 47.826

Alignment block:

US-09-769-787-162 x CNS07HFA

Align seg 1/1 to: CNS07HFA from: 1 to: 557

```
118 LeuGIuLeuMetGlnIleAlaGlnIleAlaAspArgLysProAspLysLe 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 CTGCAGGCGTGCAGAGCTTCCGGTTACGGCCATCGCAAAACGACGTA 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 uSerGlyGlnGlnGlnAlaArgValThrIleuAlaCysAlaLeuAlaVal 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112 GTCTGGCGGTGACAGACAGCTGTCGCCGTGGCGCGCGCTGGTGATCC 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 sProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLys 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 GACCCCAAGTCTGCTGTGATGAGCCGTCTGCGTGATATAAAT 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 LeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 CTGCGTCTCTCATGACAGCTTGAGTTGAAGCATTCACACGACGTTGG 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 yIleThrThrValTyrValThrHisAspGlnGlnGluAlaMetAlaLys 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 CCGTACACACCGCTCTGTCGACACAGCATGACGAGCGCAAGCGTCGAC 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 eAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArg 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
312 GCGATCGCTGGTGGTGGCGGACATGTACGCCAGATCCGACAG 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 ProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
362 CCGGATGAGATTTATTCGTCGTCGCCGAGATCCCTTTGCGCGGTTTGT 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 eGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
412 AGCGATGTGAATATTTCGCGGTCTGATACGTCACACTCC...GATAGGG 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 lArgTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeu 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
459 CGCGCATGTGATCTGGCGGTGAACGTGTACGCTGCCACGCG... 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 GlnValGluGluAlaIleHisValSerIle 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503 .....TAACTGTCCATGCTCGAT 523
```

seq_name: gb_gss:BH614583

seq_documentation_block:

LOCUS BH614583

DEFINITION 288ED9 Subclones from overlapping BAC clones spanning the hrp

cluster of *Erwinia carotovora* subsp. *atroseptica* *Pectobacterium*

carotovorum subsp. *atrosepticum* genomic, DNA sequence.

ACCESSION BH614583

VERSION BH614583.1 GI:18078383

KEYWORDS GSS.

SOURCE *Pectobacterium carotovorum* subsp. *atrosepticum*.

ORGANISM *Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium*.

REFERENCE 1 (bases 1 to 569)

AUTHORS Bell, K.S., Avrova, A.O., Holeva, M.C., Cardle, L., Morris, W., De Jong

TITLE 'W', Toth, I.K., Waugh, R., Bryan, G.J., and Birch, P.R.J.

JOURNAL Sample sequencing of a selected region of the genome of *Erwinia*

carotovora subsp. *atroseptica* reveals a candidate phytopathogenicity

genes and allows comparison with *Escherichia coli*

Unpublished (2002)

COMMENT Contact: Bell KS

Scottish Crop Research Institute

Invergowrie, Dundee, United Kingdom, DD2 5DA

Email: kbells@scri.sari.ac.uk

Class: BAC subclone.

FEATURES

source

Location/Qualifiers

1..569

/organism="Pectobacterium carotovorum subsp. atrosepticum"

/strain="SCRI039"

/db_xref="taxon:29471"

/clone_lib="Subclones from overlapping BAC clones spanning

the hrp cluster of *Erwinia carotovora* subsp. *atroseptica*"

/note="Vector: pGEM 3zf; DNA from BAC clones Eca2B8 and

Eca1C22 was rebligated and cloned into pGEM 3zf and

sequenced with SP6 or T7 primers"

alignment_scores:

Quality: 365.50 Length: 158
Ratio: 2.948 Gaps: 1
Percent Similarity: 78.481 Percent Identity: 45.570

alignment_block:

US-09-769-787-162 x BH614583

Align seg 1/1 to: BH614583 from: 1 to: 569

```
8 AsnAlaLysLysIleTyrHis...AspValProValIleGlnAsnLeuAs 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 AACGTCACAGAACCTACGACGCGGACCGCTGATGTCAAAAACCTGAA 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
23 nIleThrIleProLysGlySerLeuPheThrLeuLeuGlyAlaSerGlyC 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 TCGATATATTTCGTAAGGTGAATTTCTGACCCGCTGGGCGCGCGCT 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 ySgLYLysThrThrIleuAlaArgMetIleAlaGlyPheAsnSerIleGlu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 CGGGAAGAAACACCACTTTGATGATGCTCGCGGTTTGAACCCACAG 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 GlyGlyGluPheTyrPheAspThrLysIleAsnAspMetGluProse 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 CAGGAGAAATCTGCTGCGGATGACACCGTTGATCATCTCCGCCGCA 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 rLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIlePheProHis 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 TCAGCGCGACATCGCATGTTGTTTCAAGACTACGACGATTTCCACACA 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90 euThrValArgAspAsnValAlaPheGlyLeuMetClnLysLysValPro 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
346 TGACTGTGGCGGAGAACTCGCGGTTTCCGCTGTCATCCGCGCTTAAC 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 LysGluGluLeuIleGlnGlnIlePheAsnLysTyrLeuGluLeuMetGln 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 CGGCTGATATCAAGAGAAAGTCGATCGCTGATGCGCTAAGCT 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 eAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyGlnGlnGln 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 GACTACCTAGCGGATCGCTACCCGACAGATGTCGGCGGCGACACAG 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 lArgValThrLeuAlaCysAlaLeuAlaValAsnProSerValLeuLeu 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
496 AGGCTGTGGCGGTGCTCCGCGCTGGTGTTCGAACCGAAGCTGTGCTG 545
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 MetAspGluProLeuSerAsnLeu 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
546 ATGGATGAACCACTGGGTGCGCTG 569
```

seq_name: gb_gss:U80962

seq_documentation_block:

LOCUS U80962

DEFINITION U80962 Pseudomonas aeruginosa PAO1 (ATCC15692) Pseudomonas

aeruginosa genomic, DNA sequence.

ACCESSION U80962

VERSION U80962.1 GI:1778794

KEYWORDS GSS.

804 bp DNA linear GSS 01-MAR-2001


```

145 acysalalealavalasnproservalleuleumetasp1uproleus 162
      |||..... |||..... |||..... |||..... |||.....
206 GCGCGCATGTCGCCAACCAGAGCTTCTCTTGAGACCGCGCT 255
      |||..... |||..... |||..... |||..... |||.....
162 etrasleuualalylsleuarleuaspmetarginalaileatgglu 178
      |||..... |||..... |||..... |||..... |||.....
256 CGAACCTCGATGCCAAACCTGCTACGAGATGCGACCGAAATCAAGAG 305
      |||..... |||..... |||..... |||..... |||.....
179 lileglhileuvalglylethrthrvallthrthraspgingl 195
      |||..... |||..... |||..... |||..... |||.....
306 CTGACGCCAAGGTGCACTACGCTGTATGTGACCCAGATGAGT 355
      |||..... |||..... |||..... |||..... |||.....
195 ugualametalalieserasp1n.llealavalmetlysasp1y 210
      |||..... |||..... |||..... |||..... |||.....
356 CGAGGGGATGACGCTGGCCGACGATTTCTCATCATGCGCAGCGC 402

seq_name: gb_est2:BE552482

seq_documentation_block:
LOCUS      BE552482                639 bp    mRNA    linear    EST 10-AUG-2000
DEFINITION SMOVL2CASI2F01SK Onchocerca volvulus L2 larvae cDNA (SAM98MIM-OVL2)
ACCESSION  BE552482
VERSION    BE552482.1 GI:9794174
KEYWORDS   EST.
SOURCE     Onchocerca volvulus.
            Onchocerca volvulus.
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE  1 (bases 1 to 639)
AUTHORS   Williams, S.A.
TITLE     Genes expressed in L2 larvae of Onchocerca volvulus
JOURNAL   Unpublished (1999)
COMMENT   Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pbluescript SK.
            Location/Qualifiers
                location=Qualifiers
                1..639
                /organism="Onchocerca volvulus"
                /db_xref="taxon:6282"
                /clone="SMOVL2CASI2F01"
                /clone_lib="Onchocerca volvulus L2 larvae cDNA
                (SAM98MIM-OVL2)"
                /dev_stage="L2"
                /lab_host="XLI-Blue MRF"
                /note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
                Xho I; Filarial nematode parasite of humans. mRNA was
                prepared from approximately 9,000 L2s isolated from
                infected mosquitoes from Kumba, Cameroon and converted to
                double-stranded cDNA using reverse transcriptase and
                oligo(dT) followed by RNase H and DNA pol I. The library
                has 7.3 x 10E4 independent recombinants and the average
                insert size is approximately 1kb. The library was
                constructed by Michelle Lizotte-Waniewski. The library is
                available from Dr.S.A.Williams, email: genome@smith.edu."
BASE COUNT      145 a      156 c      189 g      149 t
ORIGIN
alignment_scores:
    Quality: 323.00      Length: 205
    Ratio: 2.259      Gaps: 2
Percent Similarity: 69.756      Percent Identity: 31.220

alignment_block:
US-09-769-787-162 x BE552482

```

```

Align seg 1/1 to: BE552482 from: 1 to: 639

22 Leuasnillethrileprolysglyserleuphethrleuenglalase 38
      |||..... |||..... |||..... |||..... |||.....
23 ATTAAATGAACGCTGCGCCCAATGATCATCGTATCTTTGGCGCTC 72
      |||..... |||..... |||..... |||..... |||.....
38 rglcysglylysthrthrleuaurgmctilealaglypheasiseri 55
      |||..... |||..... |||..... |||..... |||.....
73 CGGTGCCGGAATAACTCGCTGATTAAAGCCATCACTGAGCTACCGCC 122
      |||..... |||..... |||..... |||..... |||.....
55 leglulglygluphetrphaspasphrlylsileasnasmetglu 71
      |||..... |||..... |||..... |||..... |||.....
123 CGCAAAAAGGCGGATTGCTCAATGGCGCGGCTACTAATGATGCCGA 172
      |||..... |||..... |||..... |||..... |||.....
72 .....Proserlysrargasnilleglymetvalphegi 82
      |||..... |||..... |||..... |||..... |||.....
173 AAAGTATCTGCTGACGCGGAAAGCGCGCTGCTATCTTTTTC 222
      |||..... |||..... |||..... |||..... |||.....
82 nasnlyralailepheprohisteuthrvalarfpasnasvalalpheg 99
      |||..... |||..... |||..... |||..... |||.....
223 GGATGCGCGGCTGTCCCGCATTTACAAGTGCCTGGCATCTCGCTACG 272
      |||..... |||..... |||..... |||..... |||.....
99 lyeumetglnlyslsvalprolysglulgluileuglthrthrasn 115
      |||..... |||..... |||..... |||..... |||.....
273 GCATG.....TCGAAAAGTATGTCGATCGATCGATCGAT 304
      |||..... |||..... |||..... |||..... |||.....
116 lystylleugluleumetglnillealaglntrlyalalasparglysproas 132
      |||..... |||..... |||..... |||..... |||.....
305 AACCTGCTGGCGCTTTAGGCAATTAACCGCTTCTTGACCGTTTACCAGG 354
      |||..... |||..... |||..... |||..... |||.....
132 plylsleuserglylglnlnglnargvallthrleualacysalaleua 149
      |||..... |||..... |||..... |||..... |||.....
355 CAGCTTCTCCGGAGGCGAAACACGCGCTGGCATTTGGCGGCTTTGC 404
      |||..... |||..... |||..... |||..... |||.....
149 lavalasnproservalleuleumetasp1uproleuserasleu 165
      |||..... |||..... |||..... |||..... |||.....
405 TGACAGACCGGAAATTCGTCTGTGATGATGACCGCTGCGCTACGTGAT 454
      |||..... |||..... |||..... |||..... |||.....
166 Alalysleuarlyleuaspmetarginalailearglulleghisgl 182
      |||..... |||..... |||..... |||..... |||.....
455 ATTCCGCGTAACGCGAACGTGTCCTTATCTCAACGCGTGACACGCGA 504
      |||..... |||..... |||..... |||..... |||.....
182 uvalglylethrthrvaltyrvalthrthraspglnlglalalame 199
      |||..... |||..... |||..... |||..... |||.....
505 AATCAACATTCGATGTTGATGTACACCTTGGCTGGATGAATCTCC 554
      |||..... |||..... |||..... |||..... |||.....
199 lalieserasp1nillealavalmetlysasp1yallileglnlile 215
      |||..... |||..... |||..... |||..... |||.....
555 ATCTGGCAGACAGATGATGTACTGTGAAAACGCTGTAAGACCTTTC 604
      |||..... |||..... |||..... |||..... |||.....
216 GLYARGPROLYSGLU 220
      |||..... |||..... |||..... |||..... |||.....
605 GCGCGCTGGAGGAAG 619

seq_name: gb_gss:AF029490

seq_documentation_block:
LOCUS      AF029490                1340 bp    DNA    linear    GSS 29-AUG-2000
DEFINITION AF029490 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 593-T7, DNA sequence.
ACCESSION  AF029490
VERSION    AF029490.1 GI:2571020
KEYWORDS   GSS.
SOURCE     Salmonella typhimurium.
            Salmonella typhimurium.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
            1 (bases 1 to 1340)
REFERENCE  Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.
            Sample sequencing of a Salmonella typhimurium LT2 lambda library:
            comparison to the Escherichia coli K12 genome
            JOURNAL   FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
            MEDLINE   99243757

```

COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@fscs.sdu.edu
Class: shotgun

FEATURES
source Location/Qualifiers
1..1340
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone_11b="593-T7"
/clone_11b="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using LA-Cor
sequencer"

BASE COUNT 308 a 382 c 293 g 291 t 66 others

ORIGIN

alignment_scores:
Quality: 320.50 Length: 227
Ratio: 2.016 Gaps: 6
Percent Similarity: 70.044 Percent Identity: 36.123

alignment_block:
US-09-769-787-162 x AF029490/rev ..

Align seg 1/1 to reverse of: AF029490 from: 1 to: 1340

```

70 MetCProserLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIle 86
1056 ATGTCCCGCCGATCAGCGCTGATTAAATGATGTTTCAGTCNTATGCGTG 1007
86 epeProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnL 103
1006 T...TTCGATATGACGTA.....GACGACATCGCTTGTCTGA 970
103 yslLysValProLysGluGluLeuIleGlnGlnThrAsnLysTyrLeuGlu 119
969 GNAAGANANTGCGAAGCGAG...ATACAGCGCGTCAATGAGATGCTGAT 923
120 leuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerG 136
922 CTGGTGCACATG...CAGAGTTGCGAAACGTAGCGCGATCCATCCG 876
136 yglGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaValAsnProS 153
875 CGNTCAGCGTACGCGCGTGGCCCTGCGCGCAGCTGCGCAAGCTCTTA 826
153 erValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArg 169
825 AGCGCTCTTGTGTGATGACCGATGGCGCGCGTGCATAAATAATATACG 776
170 leuAspMetArgGlnAlaIleArgGlnIleGlnIleGlnIleGlyIleTh 186
775 GACMGATGAGCGCTGGAAGTGTGATATCTTWAAGCGCGCTGAC 726
186 rThValTyrValThrHisAspGlnGluGluAlaMetAlaIleSerAspG 203
725 CTGGCTATGTTACGACGATCGACGAAGAAGCATGATGATGCGGGGC 676
203 lInitAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 219
675 GCATTGCCATCATGACCGTGAATAATTTGTCAAAATTTGCGAAGCGAG 626
220 GluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyAr 236
625 GAGATTTACGAGCATCCGACACCCGCTACAGCGCGAATTTATCGGATC 576
236 gThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyValTyrI 253
575 GGTAAATGTTTTCAGGCTTACTGAAGCGCGTGAAGAGAGAGCGTCTCG 526

```

253 leValPheSerAspGly.....TyrAlaLeuArgMetProAla...Leu 266
525 TCATTGACGCCGCCGCTTGGTTCATCCGTTAAAGTCGATGCGAGATGCC 476
267 AspIleValGluGlnGlnAlaIleHisValSerIleArgProGluGluPh 283
475 TCTGTCGTTGATTAATGTCGCCGCTATNTGTGGCTACGATCGGAAATAAT 426
283 eIleLysAspGluSerGlyAspIleGluGly 293
425 CATGCTGTGATGATGTCGCCGCTCGCATGCGC 395

seq_name: gb_est1:A1058144

seq_documentation_block:
LOCUS A1058144 417 bp mRNA linear EST 20-JUL-1998
DEFINITION SMOVL3CAN22H02SK Onchocerca volvulus infective larva cDNA
(SAG94WR-OvL3) Onchocerca volvulus cDNA clone SMOVL3CAN22H02 5',
mRNA sequence.
ACCESSION A1058144
VERSION A1058144.1 GI:3332010
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 417)
Williams, S.A., Lu, W., Lizotte-Waniewski, M. and Laney, S.J.
Genes expressed in Infective third stage larvae of Onchocerca
volvulus
Unpublished (1995)
CONTACT: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomes@smith.edu
Seq primer: pbluescript SK.

FEATURES
source Location/Qualifiers
1..417
/organism="Onchocerca volvulus"
/strain="Stieria Leone"
/db_xref="taxon:6282"
/clone_11b="SMOVL3CAN22H02"
/clone_11b="Onchocerca volvulus infective larva cDNA
(SAG94WR-OvL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda UniZap XR; Site_1: EcoR I; Site_2:
Xho I; Cutaneous filarial nematode parasite of humans.
mRNA was prepared from third stage infective larvae of
Onchocerca volvulus isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNAPI I. The library had 1.8 x 10⁵ independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is
available from Dr. S.A. Williams, email genomes@smith.edu."

BASE COUNT 93 a 115 c 108 g 94 t 7 others

ORIGIN

alignment_scores:
Quality: 315.00 Length: 126
Ratio: 3.029 Gaps: 1
Percent Similarity: 82.540 Percent Identity: 50.000

alignment_block:
US-09-769-787-162 x A1058144 ..

Align seg 1/1 to: A1058144 from: 1 to: 417

alignment_scores:

[illegible]

4
3
2

1
0